Funded by the European Union

ActEU Activating European Citizens´ Trust in Times of Crises and Polarisation

Reporting on Deliverable D1.4 – Delivering the teaching script to the online methods workshop on collected data and all applied methods to measure the concept of political trust and legitimacy in Europe

PROJECT	
Project number:	101094190
Project acronym:	ActEU
Project name:	Towards a new era of representative democracy - Activating European citizens' trust in times of crises and polarization
Call:	HORIZON-CL2-2022-DEMOCRACY-01
Topic:	HORIZON-CL2-2022-DEMOCRACY-01-08
Type of action:	HORIZON-RIA
Responsible service:	REA
Project starting date:	01 / 03 / 2023
Project duration:	36 months

DELIVERABLE	
Deliverable No	D4
Deliverable related No	D1.4
Work package No	WP1
Description	The script includes all relevant information taught within a series of online methods workshops organized on a regularly basis (M25-35). The workshop should be addressed to all 4 doctoral and 7 postdoctoral researchers of the project (but is also open to researchers and stakeholders all over Europe).
Lead beneficiary	USAAR
Type	R
Dissemination level	PU
Due to date	30 Sep 2025
Main author(s):	Daniela Braun, Daniel Gayo-Avello, Alex Hartland, Ann-Kathrin Reinl, Kristina Weissenbach
Contributor(s):	/
Approved by	Michael Kaeding, Liesa Döpcke (UDE)

Teaching script to the online methods workshop on collected data and all applied methods to measure the concept of political trust and legitimacy in Europe

Executive Summary or Key Conclusion or Recommendations

Summary of context: Our empirical knowledge of major questions around political legitimacy and trust as well as political polarization is surprisingly scarce and sometimes even contradictory. The reasons for this are manyfold: They range from the lack of unambiguous conceptualisations of political trust, legitimacy, and polarization to the appropriate operationalization and measurement of concepts. Numerous data sources have been produced and are available for scholarly research, but still the measurement of each of the concepts is often too superficial. Against this background, the ActEU project has produced three different interconnected types of innovative data sources. However, our aim was not only to design and publish these new data sources, but also to teach ActEU's internal researchers as well as external researchers how we collected the data and how to apply appropriate methods to measure the main concept of our project. This report seeks to systematically provide information on our teaching scripts we used in diverse teaching situations.

Added value: Our starting point was to ensure that ActEU's PhD researchers had a good understanding of how to use our new data sources to address their research questions effectively within each of our deliverables, outreach activities and scientific publications. For this, we organized together with TEPSA the ActEU Doctoral School "Democratic Frontiers: Charting Pathways for Trust and Participation in European Governance" in February 2025 in Brussels where all ActEU PhD researchers as well as external researchers have been invited to apply and to participate to get a better idea of data and methods used within the ActEU project. Accordingly, the workshop was addressed to all ActEU researchers, but was open to researchers and stakeholders all over Europe. An important feature of this doctoral school was the separate sessions labelled 'Methodological Mastery', which covered all three types of data sources: focus group discussions, experimental public opinion surveys and web scraping. Thus, the added value of this four-day workshop was to inspire scholars by showing them how to use data and methods when dealing with research questions on political legitimacy, trust, and polarisation.
Furthermore, each of the data sets and methods have been also taught in different additional settings, ranging from online workshops to hands-on-sessions together with undergraduate students as well as PhD researchers. Although it was originally intended to hold workshops exclusively in online settings, we provided in-person workshops to encourage broader discussions and reach varied research communities. The added value here was to increase our efforts to raise awareness of the ActEU data and methods among a large group of researchers, inspiring them to address research questions concerning political legitimacy, trust and polarisation, and to make use of the relevant data sources. Altogether, drawing on ActEU data sources we advance the field of research around political trust and legitimacy with regard to three main aspects: First of all, we enable researchers to measure political trust and legitimacy more broadly than it is usually done (see also Deliverable D1.3). Second, we enable researchers to investigate not only one single level of polity (e.g., the national level or the EU level only), as it is usually done when dealing with political trust and legitimacy. Instead, we take into account four levels of governance within the EU multi-level system, namely the local, the regional, the national and the EU level. Third, we consider additionally the degree of societal polarization over three currently heated policy fields – immigration,

climate change, and gender inequality. This allows us to get a better understanding on how societal polarization affects political trust and legitimacy.

Summary of Deliverable D1.4: One important aim within the ActEU project was to provide a new and more original empirical infrastructure based on an innovative combination of methods and newly collected quantitative and qualitative empirical data – the ActEU focus groups (as described in Deliverable D1.1), the ActEU web scraping data (as described in Deliverable D1.2), and the ActEU survey data (as described in Deliverable D1.3). However, not only data production and making the produced data publicly available to other researchers is an important task in scientific projects, but also to teach and train other researchers on how to use the data and methods as well as to raise additional awareness of the ActEU data and methods. Therefore, Deliverable 1.4 includes information on our different types of teaching scripts we used in diverse teaching situations organized on a regularly basis.

Key conclusions: This task was not intended to produce empirical findings which can be used to derive specific policy recommendations for policymakers and civil society, but to teach and train internal as well as external researchers interested in the topic. In fact, teaching and training are highly important for the overall impact of the project, as they make the data and methods visible to a wider community. First, the effort we have put into designing and collecting our new data sources will not only benefit internal ActEU researchers when they are writing their deliverables and scientific publications, but also the wider scientific community who have been inspired in our workshops to deal with research questions concerning political legitimacy, trust and polarisation, and to make use of the relevant data sources. Secondly, this report can be efficiently used by other professors and teachers to educate and train students as well as PhD researchers in this field. Finally, we have one explicit policy recommendation for EU-level institutions that produce their own data sets (such as Eurobarometer): the European Parliament and the European Commission should teach students, young researchers, journalists and other stakeholders how to use their data to increase its visibility and justify the money spent on collecting it.

Content

Figures/Tables

1 Introduction

The ActEU project has six key objectives which enable all members of the ActEU consortium to effectively counteract current problems of legitimacy in European representative democracies:

1) Providing an innovative conceptual framework as well as an original empirical infrastructure based on new quantitative and qualitative empirical data (focus groups, experimental surveys, web scraping) and an innovative combination thereof to study political trust, legitimacy and representation in polarizing times in the European multi-level system.
2) Mapping and investigating the issue of legitimacy of European representative democracies via a triangular approach focusing on political attitudes, political behavior and political representation.
3) Identifying a set of clear-cut factors to counteract the problem of decreasing political trust, legitimacy and representation in European democracies for immediate intervention and in the long term.
4) Analysing the context sensitivity of solutions for engendering trustworthy institutions as well as a less divided society in Europe. We will do so in terms of both level of polity (local, regional, national, EU) and three of the most polarizing policy fields (migration, environment, gender inequality).
5) Developing a toolbox of remedial actions including two toolkits for (1) European, national, regional and local policymakers, and (2) civil society and the educational sector to (re-)activate citizens and to enhance trust in and legitimacy of representative democracy. For the co-creation of the toolbox, we will systematically cooperate with and get input from civil society actors and political stakeholders as well as (young) citizens and the broader public.
6) Communicating and disseminating ActEU findings across three different target groups (the younger generations of citizens, policymakers and stakeholders, and academics) in multifaceted and innovative ways (educational cartoons "Cartooning for democracy", podcasts, videos, blogs, policy briefs, reports, and others) and building up an ActEU Civil Society Network.

This deliverable deals with the first four objectives, providing information on how to train scholars to use the original empirical infrastructure, i.e. appropriate and innovative methods to measure the concept of political trust and legitimacy in Europe, developed in the ActEU project. More specifically, this report provides detailed information on the various teaching scripts used in a series of workshops, which can be adopted by other researchers in the field.

Why is this necessary? Although there is a rich literature on political trust and legitimacy as well as political polarization in Europe, our empirical knowledge of major questions around legitimacy, trust and polarization is surprisingly scarce and sometimes even contradictory. The reasons for this are manyfold: They range from the lack of unambiguous conceptualisations of political trust, legitimacy, and polarization to the appropriate operationalization and measurement of concepts. Accordingly, numerous data sources have been produced and are available for scholarly research, but still the measurement of each of the concepts is often too superficial. Against this background, the ActEU project has produced three different types of new data sources to appropriately measure political trust, legitimacy, and polarization.

One important aim within the ActEU project was to provide a new and more original empirical infrastructure based on an innovative combination of methods and newly collected quantitative and qualitative empirical data – the ActEU focus groups (as described in Deliverable D1.1), the ActEU web scraping data (as described in Deliverable D1.2), and the ActEU survey data (as described in Deliverable D1.3). Since not only data production and making the produced data publicly available to other researchers is an important task in scientific projects to make an impact, but also to teach and train other researchers on how to use the data and methods as well as to raise additional awareness of the ActEU data and methods, Deliverable 1.4 includes information on our different types of teaching scripts we used in diverse teaching situations organized on a regularly basis. This report thus seeks to systematically provide information on our different types of teaching scripts we used in diverse teaching situations.

What did we do? Our starting point was to ensure that ActEU's PhD researchers had a good understanding of how to use our new data sources to address their research questions effectively within each of our deliverables, outreach activities and scientific publications. For this, we organized together with TEPSA the ActEU Doctoral School "Democratic Frontiers: Charting Pathways for Trust and Participation in European Governance" in February in Brussels where all ActEU PhD researchers as well as external researchers have been invited to apply and to participate to get a better idea of data and methods used within the ActEU project. The workshop was addressed to all ActEU researchers, but was also open to researchers and stakeholders all over Europe. An important feature of this doctoral school was the separate sessions labelled 'Methodological Mastery', which covered all three types of data sources: focus group discussions, experimental public opinion surveys and web scraping. Thus, the added value of this four-day workshop was to inspire scholars by showing them how to use data and methods when dealing with research questions on political legitimacy, trust, and polarisation. Furthermore, each of the data sets and methods have been also taught in different additional settings (ranging from online workshops to hands-on-sessions together with undergraduate students as well as PhD researchers). The added value here was to increase our efforts to raise awareness of the ActEU data and methods among a large group of researchers, inspiring them to address research questions concerning political legitimacy, trust and polarisation, and to make use of the relevant data sources.

The remainder of this report is structured as follows: In Section 2, we provide information on the specific events where we have trained other researchers drawing on the public opinion survey as well as the teaching script itself. In the subsequent two sections, we provide the same type of information (training events and teaching script) for the data from webscraping (Section 3) and from focus group discussions (Section 4).

2 Public Opinion Survey (Ann-Kathrin Reinl & Daniela Braun)

The goal of this chapter is twofold. Firstly, we provide information on the occasions on which we taught people how to use the ActEU Public Opinion Survey. Secondly, we have designed the report in a way that it can easily be used by others who wish to deliver an introductory course on the ActEU public opinion experimental survey.

2.1 Training events

ActEU Doctoral School: Together with TEPSA, we organized the ActEU Doctoral School "Democratic Frontiers: Charting Pathways for Trust and Participation in European Governance" in February 2025 in Brussels where all ActEU PhD researchers as well as external researchers have been invited to apply and to participate to get a better idea of data and methods used within the ActEU project, including the public opinion survey. The workshop was addressed to all ActEU researchers, but was also open to researchers and stakeholders all over Europe. Within one dedicated session of the four-day workshop, participants learned about the following topics around the survey:

- Basic information on the survey design and where to find more information on the design, translation, and implementation of the survey
- type of questions/topics covered in the survey
- survey innovation on political trust
- survey innovation on affective issue polarization
- experimental elements of the survey
- key lessons for future data collection
- examples from current research
- access to the public opinion survey
- hands-on-session

Bachelor seminar at Saarland University: The data set has been also taught in the context of a seminar on "Political Trust and Legitimacy in Times of Polarization" with undergraduate students at Saarland University (by Daniela Braun) in the winter term 2025/2026. While the overall seminar was aimed to provide students with relevant information on the concept of political trust, legitimacy, and polarization via reading and discussing the research literature, we also dealt with issues of measurement in this context. Accordingly, the following topics around the survey have been discussed in the appropriate detail:

- Basic information on the survey design and where to find more information on the design, translation, and implementation of the survey
- type of questions/topics covered in the survey
- survey innovation on political trust
- survey innovation on affective issue polarization

Summer Institute in Computational Social Science (SICSS): The data set has been also taught in the context of the SICSS-Saarbrücken in September 2025 with (under)graduate students and young researchers at Saarland University (by Daniela Braun and Rosa Navarrete). Our seminar entitled "Computing of Society: Why It Matters and How It Can Be Done" dealt with the fact that societies have been profoundly reshaped both the practice and the study of politics in times of digital transformation. In this context, we explained why computational

approaches are becoming indispensable for understanding the political world today and how they can be meaningfully applied. At the same time, computational approaches in terms of both data collection and the application of methods, need to be always developed with an eye to standard/traditional approaches of social scientific research. In line with the latter, the ActEU survey design was chosen as an example for the traditional way of studying current research questions in social sciences. Accordingly, the following topics around the survey have been discussed in the appropriate detail:

- Basic information on the survey design and where to find more information on the design, translation, and implementation of the survey
- type of questions/topics covered in the survey
- survey innovation on political trust
- survey innovation on affective issue polarization

2.2 Teaching script

This section will suggest ways in which the ActEU Public Opinion Survey can be used for training purposes. It is written as a teaching script designed to be used by others who would like to teach an introductory course on the ActEU experimental survey. The following text Is enriched by examples from the presentation slides where necessary. The entire presentation is attached to the document in Appendix A.

Course details and participation requirements: Participants will gain insights into the survey's innovations, explore initial findings, and learn how to access and use the data. The goal is for the ActEU public opinion survey to be widely used by researchers and other audiences, facilitating data analysis. The session is designed for a maximum of 20 participants and will last 90 minutes. Each participant is required to bring their own mobile phone and laptop for the hands-on parts of the session. The course will be taught in English. The course itself aims to introduce a public opinion survey dataset collected in the summer of 2024 across 10 European Union countries. The survey includes both traditional questions and two survey experiments. Participants will gain insights into the survey's innovations, explore initial findings, and learn how to access and use the data. A basic understanding of quantitative survey data is required to participate in the course. While participants will become familiar with a new dataset and its innovations, the course does not provide a general introduction to working with survey data. Additionally, participants should be proficient in using statistical software to access and analyse the data during the practical session.

Teaching methods: A wide range of teaching methods will be applied throughout the course. The participants will begin with an interactive online quiz, in which they will be asked questions related to the findings of the ActEU survey. Afterwards, the dataset and its innovations will be introduced. This will be followed by a hands-on session in which participants will independently download the data and work on assignments in small groups. The course will thus combine online activities, a lecture, and group work, with ample time for questions and discussions. A supportive and inclusive environment will be fostered, ensuring a positive learning atmosphere.

Introduction of the course: The course will begin with a 10-minute quiz designed via an online survey tool (we used www.surveymonkey.com) to engage participants right from the start. Accessible via a QR code, the online quiz will highlight key findings from the public opinion survey, setting the stage for the session. All insights are based on unweighted, pooled cross-country frequencies.

The first question asks: "Which level of political parliament within the EU do people trust the most?" Participants can choose from four answer options:

1. Local (Q38_1)
2. Regional (Q38_2)
3. National (Q38_3)
4. European (Q38_10)

According to the public opinion survey, the correct answer is "local".

The second question focuses on public preferences for further EU integration: "In which of the following policy areas would citizens most like to see deeper EU involvement?" The answer options are:

1. Border control and the management of immigration (Q24_1)
2. Climate change mitigation (Q24_2)
3. Equality between men and women (Q24_3)

The survey results indicate that "equality between men and women" is the most preferred area.

The third question asks: "When following news about recent political debates, which level of parliament causes the greatest concern among citizens?" The available options are:

1. Regional (Q41_1_C)
2. National (Q41_2_C)
3. European (Q41_1_C)

Based on the survey findings, "national" was the most frequently chosen answer.

Finally, the fourth question explores affective issue polarization: "In which policy area is resentment toward out-groups most pronounced?" The options are:

1. Climate change (Q43, Q44_1, Q44_2)
2. Feminist ideals (Q45, Q46_1, Q46_2)
3. Immigration (Q47, Q48_1, Q48_2)

The public opinion survey showed that "feminist ideals" generated the highest levels of polarization.

Figure 1 Political attitude towards different levels of polity (quiz example)

1. Which political parliament in the EU do people trust the most?

- ◯ Local
- ◯ Regional
- ◯ National
- ◯ European

1. Which political parliament in the EU do people trust the most?

- ◉ **Local**
- ◯ Regional
- ◯ National
- ◯ European

1. Which political parliament in the EU do people trust the most?

- ✓ **Local**
- ◯ Regional
- ◯ National
- ◯ European

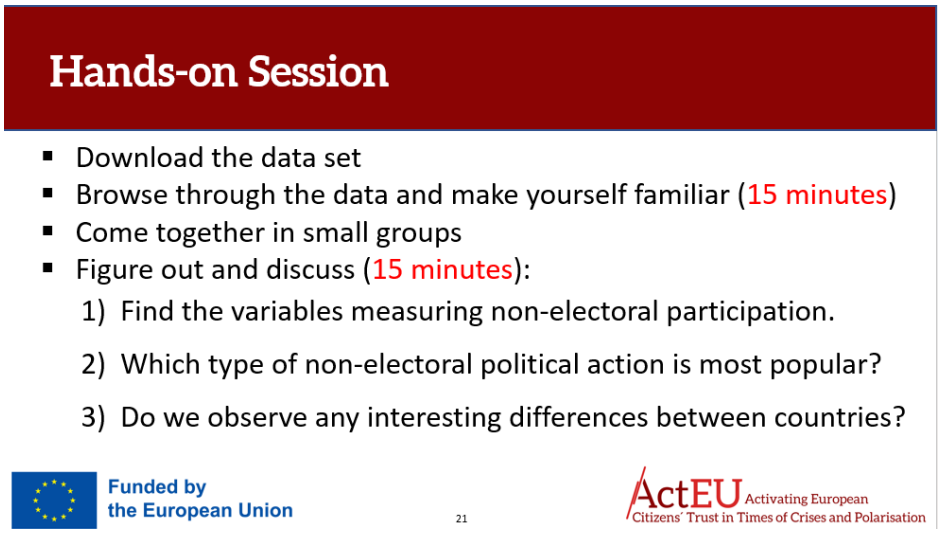
1/1 point

Lecture component: The quiz will be followed by a 35-minute lecture providing an overview of the survey, including its scope, time span, and key topics. The presenter will highlight both methodological and content-related innovations. The survey introduces original items exploring various dimensions of political trust and examines affective issue polarization on three key topics: gender, environment, and migration. Additionally, it incorporates two innovative survey experiments—a vignette experiment analysing the characteristics of political actions that encourage participation, and a conjoint experiment exploring how politicians' attributes influence public support. After discussing these innovations, the lecture will address key lessons for future data collection, covering both general insights for survey methodology and specific implications for research on political support and issue polarization. To conclude, participants will be introduced to two work-in-progress studies using ActEU survey data, providing a practical perspective on how to apply the dataset in research.

Figure 2 Political trust and legitimacy item battery

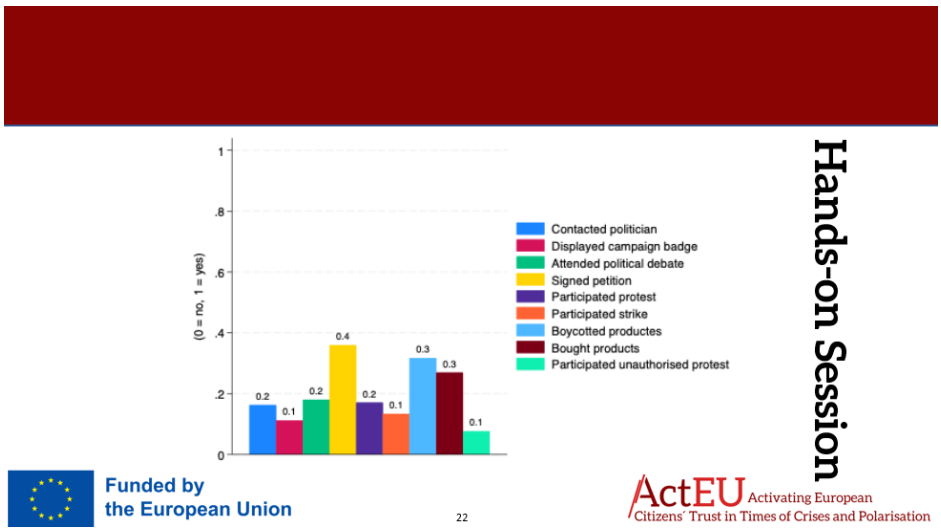
Exercise component: The session will continue with a 45-minute hands-on exercise, during which participants will learn how to access and descriptively scan the data. The values shown in parentheses in this script display the respective variable names. Before working on the hands-on exercises, participants are asked to download the data set and make themselves familiar with it. Participants may need to create an account with GESIS to access the data (max. 15 minutes will be needed for this).

Figure 3 Task I for the participants



Participants are asked to descriptively compare support for different forms of non-electoral political participation and assess their popularity (Q30a_1–Q30a_9), as well as check for any cross-country differences. Participants are granted 15 minutes to solve the question. Afterwards, the results are briefly discussed in the plenum (5 minutes). The most popular activity across countries is signing a petition. Regarding country differences, this finding holds for all countries except Germany, where boycotting certain products is most popular, and Greece, where attending an electoral or political debate attracts the most survey participants.

Figure 4 Solution to task I

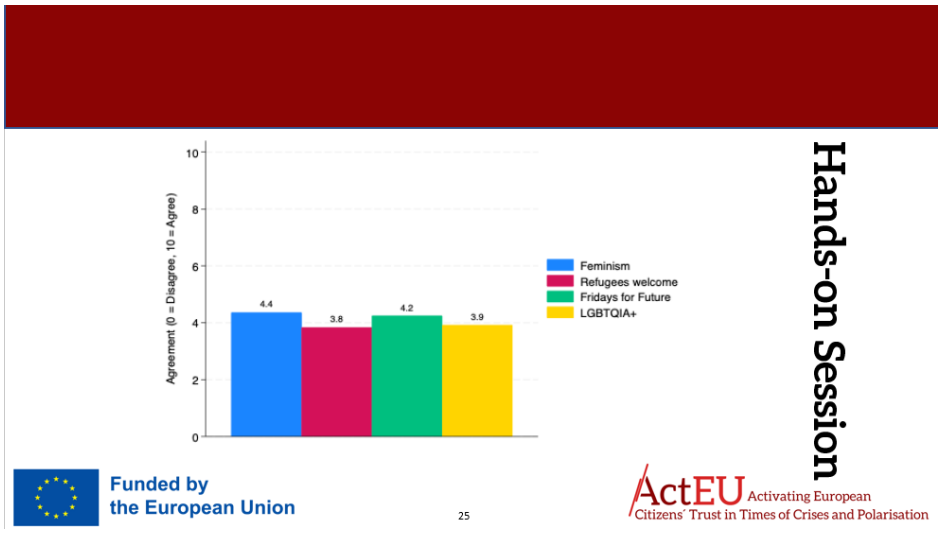


Lastly, participants will explore which flags most people would support displaying on the facades of their local parliament (Q40_1_1, Q40_2_1, Q40_3_1, Q40_4_1) and whether support varies by educational group (qu_Edu). Once again, participants are granted 15 minutes to work on the question which is followed by a discussion of the results (5 minutes). Across all respondents, displaying a feminism flag at their local council or assembly is preferred over a Refugees Welcome, Fridays for Future, or LGBTQIA+ flags. Regarding educational group differences, the highly

educated are an exception to this pattern. For this group of respondents, the Fridays for Future flag is the most popular.

Figure 5 Task II for the participants

Figure 6 Solution to task II



The workshop concludes with time for open questions.

3 Web scraping (Alex Hartland)

The goal of this chapter is twofold. Firstly, we provide information on the occasions on which we taught people how to use the ActEU web scraping data. Secondly, we have designed the report in a way that it can easily be used by others who wish to deliver an introductory course on the ActEU web scraped data.

3.1 Training events

ActEU Doctoral School: Together with TEPSA, we organized the ActEU Doctoral School "Democratic Frontiers: Charting Pathways for Trust and Participation in European Governance" in February 2025 in Brussels where all ActEU PhD researchers as well as external researchers have been invited to apply and to participate to get a better idea of data and methods used within the ActEU project, including the web scraped data. The workshop was addressed to all ActEU researchers, but was also open to researchers and stakeholders all over Europe. Within one dedicated session of the four-day workshop, participants learned about the following topics around the web scraped data:

- Defining the Task
- Who, What and Where to Scrape?
- Practical Considerations
- Details of the Content
- Future Applications of the Data

Seminar at USAAR: We organised one class as part of the Research Project (Lehrforschungsprojekt) course at the University of Saarland showing undergraduate students how they could use the ActEU web-scraped dataset to address key questions in political science related to salience, stance, and polarisation. The following topics were covered:

- Data collection
- Data cleaning
- Practical limitations
- Measuring political concepts with web-scraped dataset

Summer Institute in Computational Social Science (SICSS): As part of the SICSS workshop at the University of Saarland, graduate students covered details of methodological and practical considerations associated with using the ActEU web-scraped dataset. Following this input, students worked through hands-on examples using the data to study salience, stance, and other classification tasks via a Google Colab Notebook. The workshop included the following sections:

- Digital Texts are Digital Traces
- The Good, the Bad and the Ugly of Working with Digital Texts (mostly from Social Media)
- Targeted Collection
- Collection Methods
- Ab verbis per numeros
- Hands-on Examples

3.2 Teaching script

This section suggests ways in which the **ActEU web-scraped dataset** can be used for teaching. It is written as a script for instructors who wish to run an introductory session on digital trace data and its role in contemporary social science research. Illustrative examples from the accompanying slide deck are integrated throughout (the full set of slides is included in Appendix B).

Course overview and participation requirements

The training introduces participants to the opportunities and challenges of using digital trace data in social science research. It is organized around the **ActEU web-scraped dataset**, which contains textual material collected from multiple sources: Twitter/X, Telegram, traditional news outlets, and institutional websites.

The session runs for 90 minutes, and can be taught in English. Participants should bring a laptop for the interactive and practical parts of the course. Some prior knowledge of working with structured datasets is assumed, and basic familiarity with text analysis and coding is recommended but not essential. The focus is not on teaching generic methods, but rather on engaging critically with this dataset and understanding how it can be applied for future research.

Teaching methods

The course uses a combination of teaching methods to balance conceptual learning with hands-on practice. Short lectures provide the theoretical foundation for understanding digital traces and their role in social science research. Interactive elements, such as online quizzes and live demonstrations, are designed to keep participants engaged and to highlight real examples from the ActEU web-scraped dataset. Group-based exercises encourage collaboration and peer learning, allowing participants to experiment with the data and share different approaches. Throughout, the emphasis is on creating an inclusive and interactive classroom atmosphere where participants feel comfortable exploring new tools and discussing both the possibilities and pitfalls of working with digital text data.

Lecture Component: What We Cover in this Workshop

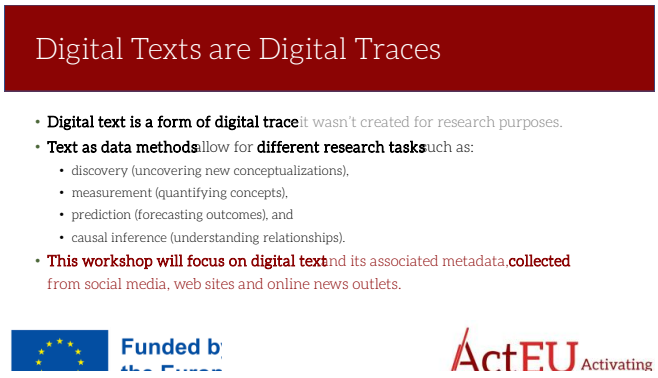
Here, the presenter will walk participants through the flow of the workshop. The outline emphasizes an understanding of digital texts as a form of digital trace and practical steps, including collection, preprocessing, and analysis. Each stage is linked to the ActEU web-scraped dataset, which will serve as the central example. By previewing the trajectory, participants will see how theoretical discussions connect to hands-on exercises and how the different components of the session build toward applied research skills.

Figure 7 Workshop Overview

Digital Texts are Digital Traces

This slide introduces the central idea that digital text is data produced for purposes other than research. Tweets, posts, articles, and website meta data are all artifacts of communication that can be systematically studied. The presenter will explain how treating text as a digital trace opens possibilities for discovery, measurement, prediction, and causal inference. The discussion underscores that the workshop is not about collecting "just text," but can be about leveraging text alongside metadata to capture context and meaning in political communication.

Figure 8 Digital texts are digital traces



Targeted Collection

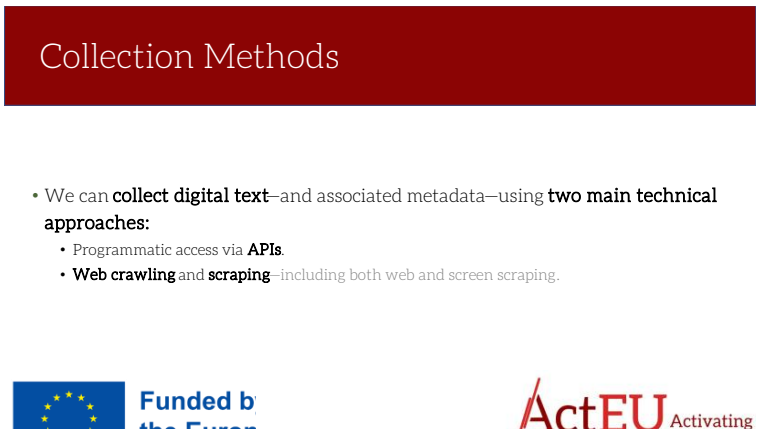
Here the focus shifts to how researchers define and narrow the scope of their data collection. The presenter will emphasize that meaningful analysis starts with careful decisions about what to include: identifying relevant platforms, actors, events, or keywords. Examples will show how researchers might focus on Telegram channels tied to political groups, institutional websites addressing policy issues, or popular social media platforms where political content is commonly shared. This part of the session is meant to make participants reflect critically on the link between their research questions and the specific traces they decide to collect.

Figure 9 Targeted collection

Collection Methods

This section introduces the technical strategies for actually obtaining digital text data. The presenter will contrast programmatic collection via APIs with web crawling and scraping, explaining the advantages and constraints of each. The ActEU dataset will be used as an example of how these methods can be combined to create a multi-source corpus. The goal here is not just to describe techniques, but to highlight the practical trade-offs researchers face: between completeness and feasibility, openness and restrictions, or automation and quality control.

Figure 10 Collection methods



Text Preprocessing

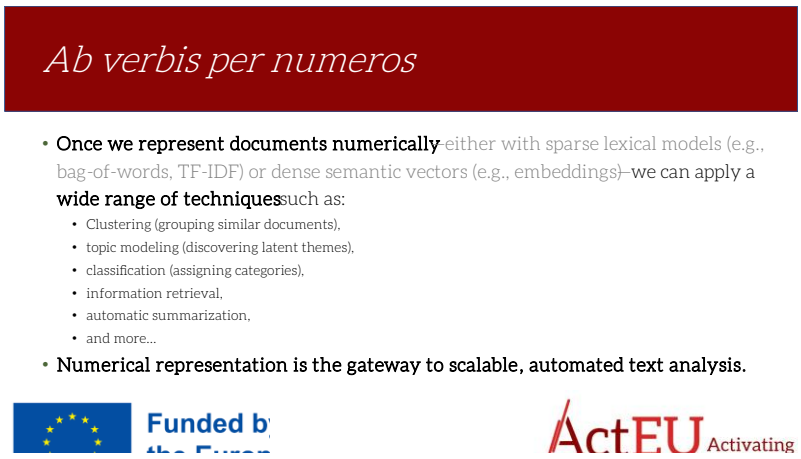
The presenter will introduce preprocessing as a bridge between raw digital text and analyzable data. Online content rarely comes in a clean, ready-to-use form. Instead, it is wrapped in formatting, advertisements, navigation menus, or other elements that are irrelevant to research. This is where "boilerplate removal" comes in: the process of stripping away extraneous material to isolate the meaningful text. Participants will learn that preprocessing is less about perfectly understanding language and more about applying structured simplifications to make complex material computationally tractable. By looking at examples, from news articles cluttered with sidebars to scraped web pages filled with code, the section demonstrates why cleaning is indispensable, and how it determines the quality and reliability of downstream analysis.

Figure 11 Text preprocessing

Ab verbis per numeros

The final conceptual step is the transformation of words into numbers. The presenter will explain how text can be represented through simple lexical models or more sophisticated semantic embeddings. Once in numerical form, text can be clustered, classified, or modeled to reveal hidden structures and patterns. This transition from qualitative meaning to quantitative representation is presented as the key gateway to large-scale automated analysis.

Figure 12 Representing text as numbers



Hands On Examples

The session concludes by pointing to hands-on exercises, where participants will explore these transformations themselves, using the ActEU web-scraped dataset to measure salience, sentiment, and polarization in political communication. The code for these examples can be found here:

https://colab.research.google.com/drive/1hOSgGYcfQ9p7ctFhA6t1cwy6J6naGjlj?usp=sharing#scrollTo=UEojMXFCcZeH

The hands-on examples cover the following details:

Introduction

This notebook accompanies the ActEU Online Methods Workshop. It demonstrates how to work with the annotated Twitter/X data released as part of the ActEU project, focusing on gender issues during the fortnight centered on November 25, 2023 (the International Day for the Elimination of Violence against Women).

In this session, participants will learn how to:

- Load and filter a sample of annotated tweets
- Produce time series for issue salience and position (stance)
- Create interactive visualizations
- Perform topic modeling using Top2Vec
- Use a large language model (LLM) to generate descriptive English labels and explanations for the discovered topics

All of this material has been designed with accessibility in mind. Code was developed using "vibe coding," so participants are encouraged not to panic — the focus is on exploration. Remember that in Google Colab, you can always click the kebab menu on a code cell and select Explain code if needed.

Step 0 – Prepare the Environment

We begin by setting up the environment and installing the Top2Vec package for topic modeling. Because Top2Vec downgrades some dependencies, a runtime restart is required.

Step 1 – Download the Data

We download the annotated tweet sample and accompanying metadata. These files include:

tweets_25nov.zip – a sample of annotated tweets from selected key accounts around November 25, 2023

ActEU-twitter-seed-with-wikidata.xlsx – metadata for key accounts, including country of origin and actor type

Step 2 – Load and Organize the Data

Next, we extract the tweets and associate each one with:

- The country of the author (from metadata)
- The actor type (e.g., politician, party, journalist, interest group)
- The publication date

This creates the core dataset for analysis.

Step 3 – Visualize Tweet Volume by Country

We generate a bar chart showing the number of tweets collected per country. This overview highlights data density and imbalances (e.g., overrepresentation of Finland and Spain, underrepresentation of Germany, Italy, and Greece).

Step 4 – Tweet Volume Over Time

We plot time series of tweet volume per country to detect spikes, drops, or anomalies. For example, a notable decline appears on November 27, likely due to scraping issues after Twitter/X API restrictions in late 2023.

Step 5 – Compute Issue Salience

We calculate salience for Climate, Gender, and Migration by normalizing confidence scores for each issue per day. This measure captures the proportion of attention allocated to each issue.

Step 6 – Interactive Visualization of Issue Salience

Using interactive charts, participants explore how issue salience varies by country over time. Peaks around November 25 are visible in some countries (e.g., EU level, France, Spain), while irregularities in others highlight data collection limitations.

Step 7 – Analyze Position (Stance)

Each tweet is annotated with stance categories (Positive, Negative, Neutral, Other). We aggregate stance data by issue, country, and date to observe how discourse leans across time and context.

Step 8 – Salience-Weighted Position Analysis

We refine the stance analysis by weighting stance values with issue salience, ensuring that both attention and position are reflected in the results.

Step 9 – Interactive Visualization of Stance Over Time

Stacked area charts show stance distributions (Positive, Negative, Neutral, Other) over time for each country. Peaks around November 25 illustrate moments of heightened attention, often accompanied by shifts in stance balance.

Step 10 – Extract Tweets on Gender Issues

We filter the dataset to retain only tweets explicitly annotated as relating to Gender. These tweets form the subset for topic modeling.

Step 11 – Sampling the Data

To ensure efficient modeling, we sample up to 1,000 tweets per country. Countries with fewer tweets are excluded. This keeps processing manageable while retaining analytical richness.

Step 12 – Topic Modeling with Top2Vec

Using Top2Vec with a multilingual embedding model, we identify latent topics within gender-related tweets. This step reveals recurring themes and clusters of discourse across countries.

Step 13 – Produce Topic Overviews

For each country, we examine the number of topics, top keywords, and representative tweets. This provides an initial, data-driven view of elite gender discourse.

Step 14 – Generate Descriptive Topic Labels with an LLM

To make topics more interpretable, we prompt an LLM with keywords and sample tweets, asking it to return concise labels and short explanations. This step bridges machine-driven topic extraction with human-friendly interpretation.

Step 15 – Create Prompts for the LLM

We structure prompts systematically so that each includes both keywords and representative tweets. This ensures consistency and clarity in the LLM outputs.

Step 16 – Send Prompts to Gemini and Parse Responses

The prompts are processed with the Gemini model, which returns JSON-formatted outputs containing a label and explanation for each topic.

Step 17 – Interactive Exploration of Topics

Participants can select countries and topics via dropdown menus to explore:

- Keywords
- Representative tweets
- LLM-generated labels and explanations

This interface makes it easy to validate and compare topic interpretations across contexts.

Conclusion

In this notebook, we demonstrated how to collect, process, and analyze political discourse on gender issues using annotated Twitter/X data from the ActEU project. Starting with raw tweets, we moved through:

- Data filtering and visualization – tweet volume and issue salience
- Stance analysis – combining position with salience to reveal dynamics in elite discourse
- Topic modeling – uncovering key themes in gender-related conversations across countries
- LLM-assisted interpretation – generating human-readable labels and explanations to enhance interpretability

This workflow shows how computational methods, interactive visualizations, and generative AI can work together to uncover insights in political communication.

At the same time, participants are reminded of the limitations: data sparsity, API restrictions, annotation uncertainty, and model biases.

Ultimately, the notebook offers a reproducible framework for analyzing digital political discourse, illustrating how traditional social science questions can be combined with modern computational and AI-driven approaches.

4 Focus group discussions (Kristina Weissenbach)

The goal of this chapter is twofold. Firstly, we provide information on the occasions on which we taught people how to use the ActEU focus grouop data. Secondly, we have designed the report in a way that it can easily be used by others who wish to deliver an introductory course on the ActEU focus group data.

4.1 Training events

ActEU Doctoral School: Together with TEPSA, we organized the ActEU Doctoral School "Democratic Frontiers: Charting Pathways for Trust and Participation in European Governance" in February 2025 in Brussels where all ActEU PhD researchers as well as external researchers have been invited to apply and to participate to get a better idea of data and methods used within the ActEU project, including the web scraped data. The workshop was addressed to all ActEU researchers, but was also open to researchers and stakeholders all over Europe. Within one dedicated session of the four-day workshop, participants learned about the following topics around the focus group data:

- How to include focus groups in (different) mixed-methods designs with a focus on exploratory instrument design mixed methods studies.
- Gaining deep knowledge about how to conceptualize and test a focus group guide, about sampling and recruiting of participants as well as the planning, organizing and moderation techniques.
- Getting an overview about different forms of collecting and analyzing focus group data.
- How to apply qualitative content analysis (QCA) and thematic analysis using theoretical concepts, categories and the coding frame.
- The instrument development function of focus group data: How can findings from a focus group data based thematic analysis help building better items of a survey (develop better wording or more comprehensive closed answers), help design a survey experiment or inform a webscraping exercise about words in context (embedded words segments) alongside a theoretical concept.
- Hands-on workshops for students to reflect their own research projects and share it with young and established scholars: Which kind of data helps finding answers to which conceptual dimension of my research question? How might focus group data complement my design? One method of data conduction and analysis – or multiple? Which design to triangulate my data?

Workshop Series at the Summer School of the ECPR Standing Group on Political Parties 2023-2025:

The ActEU Focus Group data set and ActEU methodological approach has been also taught in a methods workshop series in the context of the Summer School of the ECPR Standing Group on Political Parties "Theories, Methods and Practice Perspectives in Party Politics: Investigating Parties and Party Systems in young and established democracies" organized at the University of Duisburg-Essen, Institute for Political Science and NRW School of Governance. The program consists of a serious of lectures as well as in-depth discussions of participants' research projects and a practitioners' roundtable "Party Research meets Party Politics". Since methods and data have become more diverse in the field of party politics, the program includes hands-on sessions on research design, methods data collection and data bases, analysis, management and knowledge transfer in party politics research. The methods workshop "Working with concepts & Focus Group Data" (by Kristina Weissenbach) dealt with different questions in the different years of the

worshop series: "Political Participation in and beyond parties" (2023), "Realizing linkage in the Twenty-First century" (2024) and "When do citizens trust political parties?"(2025)

The following topics around the ActEU data set have been discussed in the appropriate detail:

- Basic information on the ActEU exploratory and sequential mixed methods design and where to find more information on conceptualization, design and implementation of the focus group data.
- Innovative data conduction methods via focus groups, like Chinese portraits or experimental settings.
- Hands-on transcription and coding phases.
- How to apply qualitative content analysis (QCA) and thematic analysis using theoretical concepts, categories and the coding frame.
- The instrument development function of focus group data: How can findings from a focus group data based thematic analysis help building better items of a survey (develop better wording or more comprehensive closed answers), help design a survey experiment or inform a webscraping exercise about words in context (embedded words segments) alongside a theoretical concept.
- Hands-on workshops for students to reflect their own research projects and share it with young and established scholars: Which kind of data helps finding answers to which conceptual dimension of my research question? How might focus group data complement my design? One method of data conduction and analysis – or multiple? Which design to triangulate my data

4.2 Teaching script

This section will suggest ways in which the ActEU Focus Group design and data can be used for training purposes. It is written as a teaching script designed to be used by others who would like to teach an introductory course on focus group conceptualization, data conduction and analysis. The following text Is enriched by examples from the presentation slides where necessary. The entire presentation is attached to the document in Appendix C.

Course details and participation requirements: Participants will learn how to use concepts and build conceptualizations for designs, guides and code books for focus group discussions in comparative politics. This is taught based on the ActEU focus group design and data conducted in 16 focus groups in the countries France, Germany, Greece and Czech Republic, exemplified by the research field of political participation and the role of emotions in political participation. The students will gain insights into the ActEU focus group design, screening and recruiting process and learn about different designs to triangulate data, with a focus on the ActEU exploratory designs. They will learn how to access and use the data, with a focus on how to apply qualitative content analysis (QCA) and thematic analysis using theoretical concepts, categories and the coding frame. The students will be able to reflect their own research projects and share it with young and established scholars: Which kind of data helps finding answers to which conceptual dimension of my research question? How might the ActEUfocus group data complement my design? One method of data conduction and analysis – or multiple? Which design to triangulate my data?

The goal is for the ActEU focus group transcripts to be widely used by researchers and other audiences, facilitating data analysis. The session is designed for a maximum of 20 participants and will last 90 minutes (without coding exercises) to 120 minutes (incuding coding exercises). Each participant is required to bring his or her own mobile phone and laptop for the hands-on parts of the session. Each participant is required to install a software for qualitative data analysis

(preferable MaxQDA free trial version). The course will be taught in English. A basic understanding of qualitative content analysis is required to participate in the course.

Teaching methods: A wide range of teaching methods will be applied throughout the course. In every single phase of the workshop, participants will be asked *to transfer the gained knowledge to their own research project* and to make according notes on colorfoul sticky notes / digital miro sticky notes to be shared and discussed with the group in the end. The participants will begin with an *exploratory, narrative phase* based on *prompts* (from the ActEU focus groups guide) regarding different conceptualizations of political participation (within and outside the institutions of representative democracy), in which they will be asked to describe concepts of participation on the indicator level. The narrative phase is followed by several *lecture phase* and brief *feedback rounds* where definitions and methodological, theoretical and practical benefits of focus group data are explained as well as different types of multi-method designs and the exploratory and sequential ActEU design is investigated. Afterwards, the student's research projects (the sticky notes) will be *briefly presented* and core challenges regarding their focus group designs and conceptualizations will be discussed in a *feeback-round* .

Afterwards, the practical side of focus group data conduction will be explained: How to design, organize and conduct focus groups; sampling and recruiting; The ActEU focus group conceptualization, design, sampling and guide as well as its innovations and moderation techniques will be introduced.

This will be followed by a hands-on session on how to analyse focus group data (transcripts and video or audio material) through a.) qualitative content analysis and b.) thematic analysis, in which participants will independently download the data into the software for qualitative data analysis and work on coding-assignments in small groups. The course will thus combine narrative phases, transfer phases, presentation phase, online activities, a lecture, and group work, with ample time for questions and discussions. A supportive and inclusive environment will be fostered, ensuring a positive learning atmosphere.

Required readings and preparations:

- *If coding exercise is applied, set up software and transcript-project*: Students should be informed to bring their own laptop and download a (free test) version of qualitative data analysis software (preferable the same throughout the course, e.g. MaxQDA or Atlasti.) and upload the transcript of one 90minute focus group discussion transcript *before* the class starts.
- Either physical colourful sticky notes are handed out to the participants or a digital sticky note tool (like miro.com) is being used.
- The synopsis of multi-method designs Creswell/Plano Clark (2011) is handed out to the students in print.
- Readings:
    - ActEUreport on D1.1, chapter 1,2,4 and 5
    - Braun, Virgina / Clarke, Victoria (2006) Using thematic analysis in psychology, Qualitative Research in Psychology, 3:2, 77-101, DOI: 10.1191/1478088706qp063oa.
    - Creswell, John W. / Plano Clark, Vicki L (2011/2017): Designing and Conducting Mixed Methods Research, London: Sage, 179-201.
    - Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, A. G. (2009). A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research. International Journal of Qualitative Methods, 8(3), 1-21. https://doi.org/10.1177/160940690900800301

Introduction of the course:

1. Narrative Phase (10 minutes): Being puzzled about something: Unit of analysis & Research Question. The demand side of participation

The participants will begin with an exploratory narrative phase based on prompts (from the ActEU focus groups guide) to jointly develop conceptualizations of political participation (within and outside the institutions of representative democracy), in which they will be asked to describe concepts of participation on the indicator level.

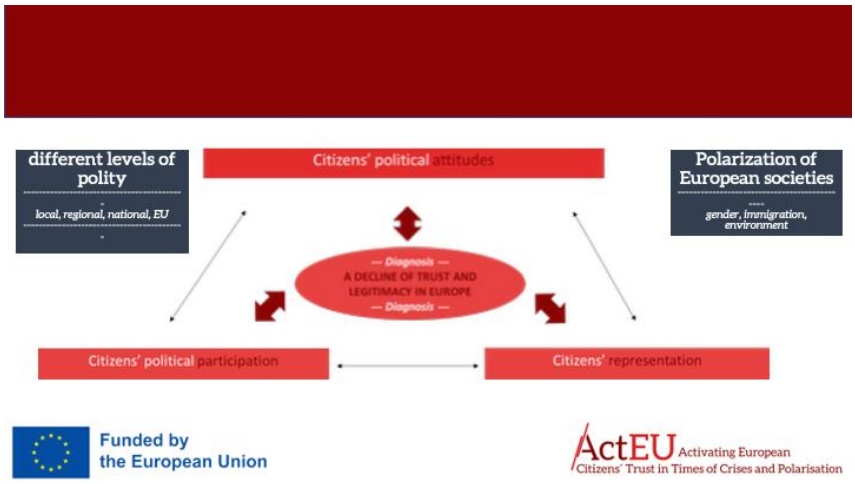
Figure 13 Prompting Political Participation inside and outside the institutions of representative democracy

Participants are asked to make their first sticky note: What are you puzzled about in your research project? Which is your main research question?

2. Lecture component: Concepts and Conceptualization (10 minutes)

In a 10 minute lecture the differences between theoretical concepts and their conceptualization on the indicator level are thaught, the ActEU theoretical concept is being introduced.

Figure 14 The ActEU theoretical concept



Participants are asked to make their second sticky note: which are the main theoretical concepts of your research project?

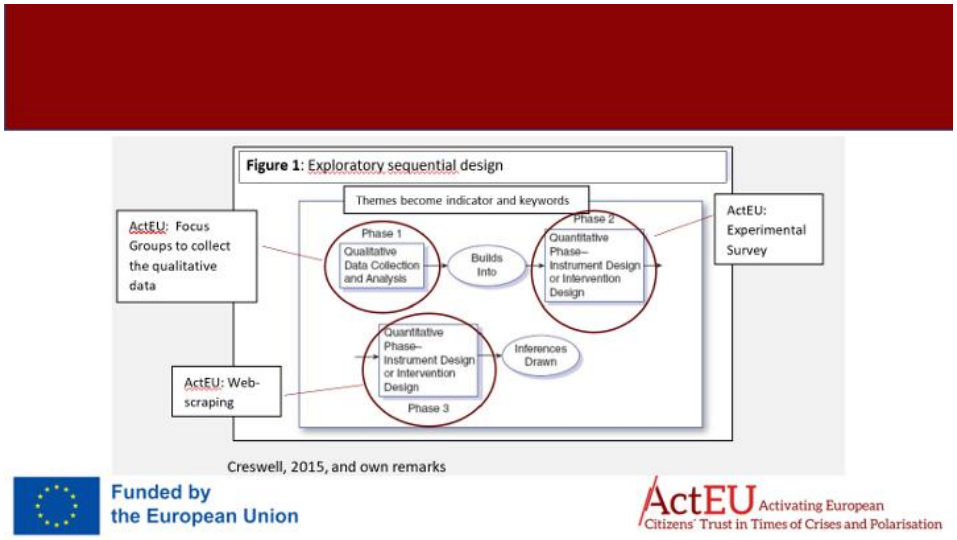
3. Lecture component: Which kind of data for our research question? Why focus groups? One method of data conduction and analysis – or multiple? Which multi-method design to triangulate your data? (15 minutes)

Definitions and methodological, theoretical and practical benefits of focus group data are explained as well as different types of multi-method designs and the exploratory and sequential ActEU design is investigated. Based on the readings of the course the typology for multi-method designs by Creswell and Plano-Clark (Creswell-Plano 2011; Creswell 2013) are introduced, the synopsis of multi-method designs is handed out to the students in print. Each design is taught in the lecture:

- Convergent parallel design
- Explanatory sequential design
- Exploratory and sequential design
- Embedded design
- Transformative design
- Mulitphase design

The exlporatory and sequential design of ActEU is being explained in the lecture component with a focus on methodological, theoretical and practical benefits and values of focus groups in multi-method designs.

Figure 15 The ActEU Exploratory sequential design - how to include focus groups in your multi-method design



Participants are asked to make their third sticky note: Is your research project based on new data or existing data? One method of data conduction or multiple? Which?

4. Transfer-Buzz group: Which multi-method design to triangulate your data – how to include the ActEU focus group data into your research design? (10 minutes)

Students are grouped in Buzz Groups (a small group work of two) and are asked to:

Please take a moment to a.) have a look and discuss the Creswell/Plano-typology for multi-method research designs and b.) collect disadvantages and advantages / the applicability to your research project desgin.

*Participants are asked to make their fourth note: **If** your research project is based on a multi-method design*

- *which data to cover which aspect or dimension of your unit of analysis?*
- *Which multi-method design to triangulate your data?*

5. Presentations: Share your sticky notes with the group! (20 minutes)

The Buzz-Groups will be followed by a presentation phase. Due to time restrictions only a third of the students, e.g. six students, are asked to present their developed sticky notes. Students' research projects will be *briefly presented* and core challenges regarding their focus group designs and conceptualizations as well as open questions from the whole group will be discussed in a joint *feeback-round.*

6. Lecture component: How to design, organize, conduct and analyze focus groups via thematic analysis (15 minutes)

In a 20 minutes lecture component students get an introduction on sampling and recruiting, moderation techniques as well as Qualitative Content Analysis (QCA) and Thematic Analysis/Constant comparison analysis. Innovations from the Act EU focus group data conduction, like the method of Chinese portraits, is introduced as well as the ActEU coding circle and examples from the ActEU MaxQDA projects. The Coding Circle can be used for a potential final *Hands on coding session.*

Figure 16 Innovative Data conduction - Chinese Portraits

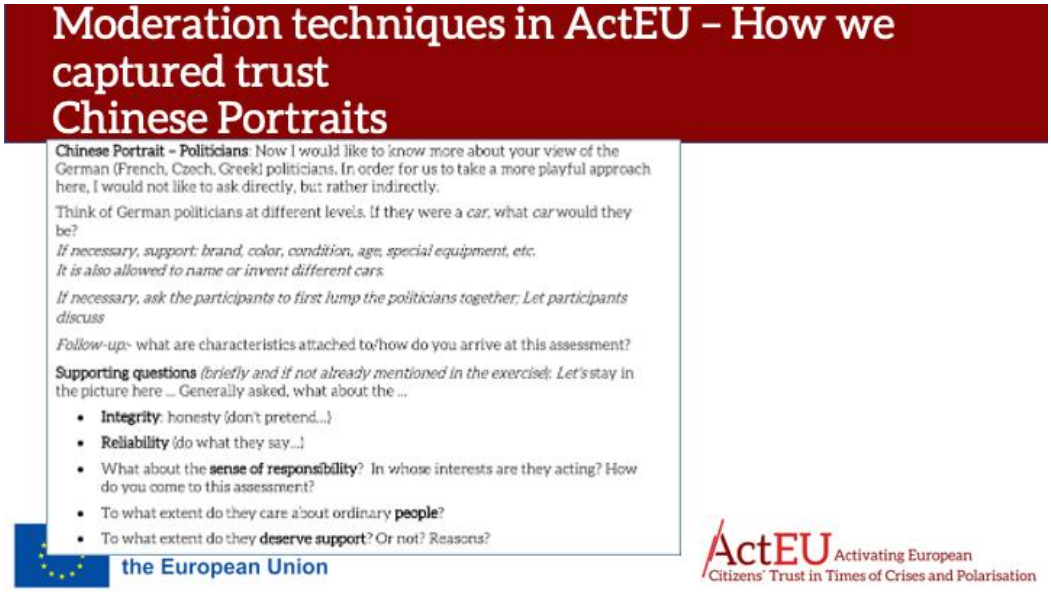
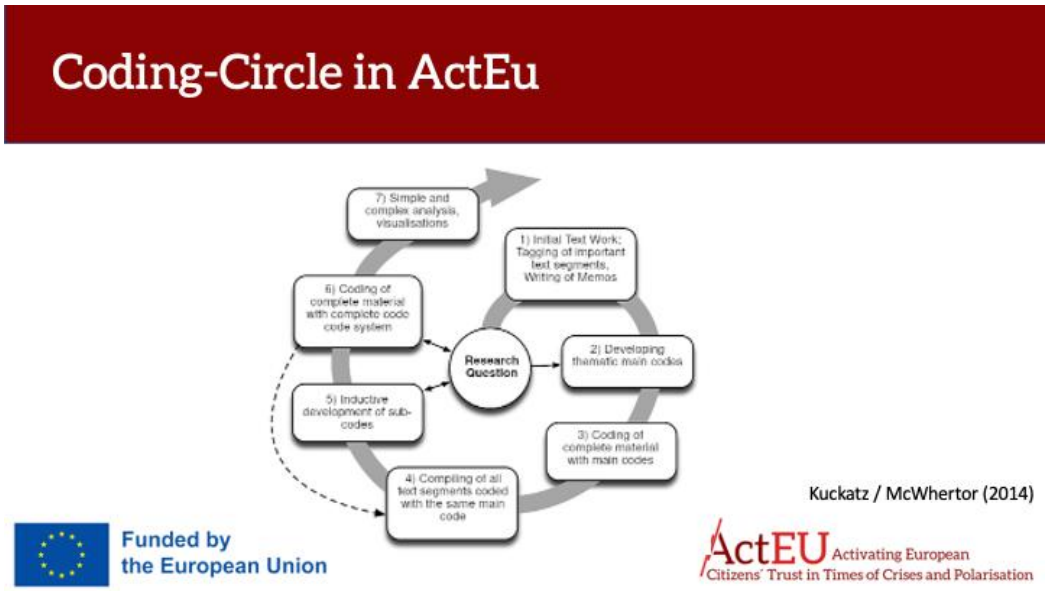


Figure 17 Coding Circle

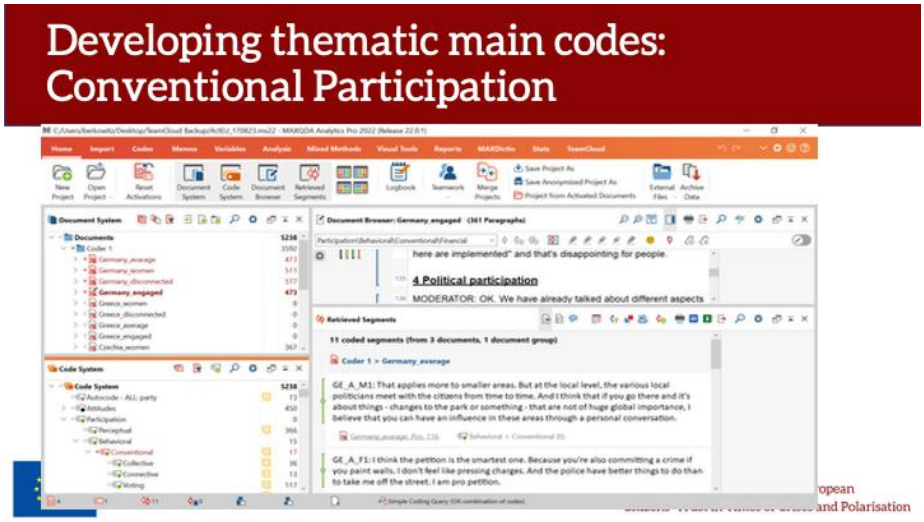


7. (Potential) Final Hands-on Phase! Try out coding alongside the coding circle with ActEU focus group data. (30 minutes)

If the long version of the class is being taught: Participants are asked to apply the coding-circle to one of the ActEU focus groups transcripts and to develop main and sub-codes in an abductive way:

- Tag important text segments according to the theoretical concept
- Develop thematic main codes according to the theoretical concept
- Code exemplary paragraphs of the ActEUfocus group transcript with main codes
- Develop thematic sub-codes inductively
- Complete the code system

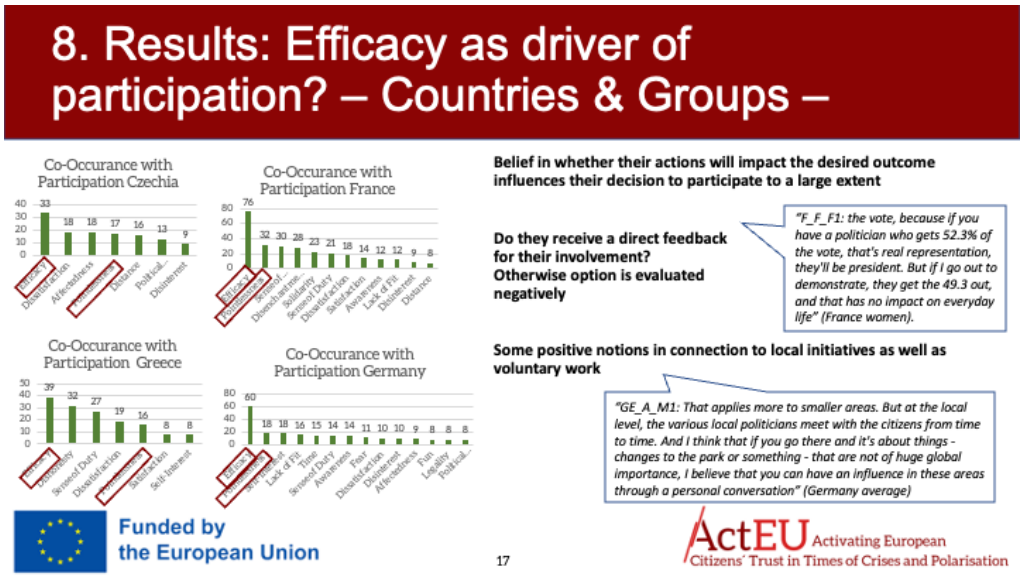
Figure 18 Developing thematic main codes



8. Closing: Insight ActEU Findings from the Focus Groups (10 minutes)

Finally, in a closing phase some of the core findings from the thematic analysis of the ActEU focus group data are presented.

Figure 19 Findings from the thematic analysis of ActEU focus group data - efficacy as driver of participation



Appendix A: Presentation slides for the public opinion survey

Methodological mastery: *Experimental Survey*

Funded by the European Union

Daniela Braun
Ann-Kathrin Reinl

ActEU Activating European Citizens' Trust in Times of Crises and Polarisat

Outline

1) Quiz Time
2) ActEU Survey Overview
3) Topics Covered
4) Survey Innovations
5) Key Lessons for Future Data Collection
6) Research Applications
7) Accessing *ActEU* Data
8) Hands-On Session

Funded by the European Union

ActEU Activating European Citizens' Trust in Times of Crises and Polaris

It's quiz time

Scan QR code

→ Please choose one option of which you think is correct and submit your answers when you're done!

Funded by the European Union

3

ActEU Activating European Citizens' Trust in Times of Crises and Polaris

ActEU Survey Overview

- Online survey; IPSOS access panel
- N respondents = 13.000
- N countries = 10
  Austria, Czech Republic, Denmark, Finland, France, Germany, Greece, Italy, Poland, Spain
- 23 minutes (median)
- Quotas according to EUROSTAT for age, gender, education (ISCED) and region (NUTS1 levels)
- Field work: 15.04.2024 – 29.05.2024

Funded by the European Union

4

ActEU Activating European Citizens' Trust in Times of Crises and Polaris

Topics Covered

1) Socio-demographics
2) General political attitudes (*LR, political interest, interpersonal tru*
3) Preferences for policy issues (*climate, gender, environment, EU*)
4) Media usage
5) Political participation
6) Representation

Funded by the European Union

5

ActEU Activating European Citizens' Trust in Times of Crises and Polaris

Survey innovation: Political trust *NEW!*

Why is political trust essential?

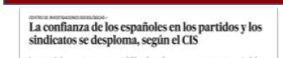
- Political trust is fundamental to representative democracy
- A trustworthy relationship between citizens and the state ensures…
  - …the functioning of democratic systems,
  - …reduces transaction costs for the system,
  - …facilitates the justification of political decisions.
- However, when we look at public debate, we see that European representative democracy appears to be under pressure, and trust seems to be declining.

Funded by the European Union

6

ActEU Activating European Citizens' Trust in Times of Crises and Polaris

Survey innovation: Political trust *NEW!*


La confianza de los españoles en los partidos y los sindicatos se desploma, según el CIS


Britons' trust in politics has fallen significantly since Covid, report finds


Zwei Drittel halten Staat für überfordert


SONDAGE EXCLUSIF – La cote de confiance d'Emmanuel Macron plombée par l'inflation

Funded by the European Union

7

ActEU Activating European Citizens' Trust in Times of Crises and Polaris

Survey innovation: Political trust *NEW!*

Why do we need a new measurement of political trust?

- **Conceptual unclarity**
  - diffuse vs specific system support (in line with Easton's theoretical framework)
  - unidimensional vs multidimensional trust (e.g. communal/value-based trust vs rational/interest-based trust; trust vs mistrust vs distrust; each concept is driven by different attitudes)
- **Lack of appropriate distinction** between the different levels of polity (local, regional, national, EU)

Funded by the European Union

8

ActEU Activating European Citizens' Trust in Times of Crises and Polaris

Survey innovation: Political trust *NEW!*

Do you agree or disagree with the following statements? (0: strongly disagree– 10: strongly agree)

1) *[Local / Regional / National / European]* MPs follow the rules.
2) *[Local / Regional / National / European MPs]* distort the facts to make policies look good.
3) *[Local / Regional / National / European MPs']* work is open and transparent.
4) *[Local / Regional / National / European MPs]* try to achieve good things.
5) *[Local / Regional / National / European MPs]* want to do their best to serve the country.
6) *[Local / Regional / National / European MPs]* understand the needs of my community.
7) I am uncertain whether or not *[local / regional / national / European MPs]* care about people like me.
8) I am unsure if *[local / regional / national / European MPs]* try to make things better or worse.
9) I am not sure how effective *[local / regional / national / European MPs]* are.
10) I am unsure whether to believe most *[local / regional / national / European MPs]*.
11) *[Local / Regional / National / European MPs]* take too long to do anything.
12) *[Local / Regional / National / European MPs]* make things worse.
13) I can have an influence on politics.
14) It doesn't matter who you vote for, politicians do whatever they want.
15) I feel that I have a pretty good understanding of the important political issues facing our country.

Survey innovation: Affective issue polarization *NEW!*

Why do we need a new measurement for issue polarization?

- Affective issue polarization in European societies is assumed to have negative effects on representative democracy
- But are our societies truly polarized?
- The academic debate is inconclusive—definitions, sources, and methods play a crucial role
- **The *ActEU* survey examines issue polarization in three key areas gender, environment, and migration**

Survey innovation: Affective issue polarization *NEW!*

Affective issue polarization: Environment I

Issues related to climate change invoke divergent feelings among people. On the one hand, there are those who believe that climate change is driven by human actions. On the other hand, there are also people who deny that climate change is driven by human actions. **Into which group would you place yourself?**

1) People who believe that climate change is driven by human actions
2) People who deny that climate change is driven by human actions
3) Neither

Affective issue polarization: Environment II

What are your feelings towards these groups of people?

(-5 very negative – +5 very positive)

1) People who believe that climate change is driven by human actions
2) People who deny that climate change is driven by human actions

Survey innovation: Vignette and Conjoint experiments *NEW!*

- Do you have any experience with survey experiments?
- Why are survey experiments necessary?
- Why can't we rely solely on traditional public opinion surveys to study European attitudes and behavior?

Survey innovation: Vignette and Conjoint experiments *NEW!*

We do not yet have a good understanding of causal mechanisms!

- To address this, we need experimental or panel data.
- How can we motivate people to politically engage?
- What are exactly the reasons for citizens' disaffection with politics?

Survey innovation: Vignette experiment *NEW!*

Vignette text:

Political decision-makers often need to take controversial decisions where some groups in society disagree with the intentions. Imagine a situation where the parliament is considering a new measure concerning the policy on *[climat change / equality between men and women / the management of immigration]* that you **disagree** with. In reaction to tha *[a group of ordinary citizens / the opposition / a celebrity]* is organising *[an online petition / a campaign to email MPs / peaceful demonstration / an occupation of the parliament building]* to show their **dissatisfaction** with the proposal.

Please indicate how likely it is that you would join this action?

0-100 vertical slider scale: 0: extremely unlikely– 100: extremely likely

How acceptable do you think this action would be?

0-100 vertical slider scale: 0: completely unacceptable – 100 completely acceptable

How likely do you think it is that decision-makers would listen to the concerns raised by the activists?

0-100 vertical slider scale: 0: extremely unlikely– 100: extremely likely

Survey innovation: Conjoint experiment *NEW!*

Conjoint introduction

There are many different reasons why people like or dislike politicians. In the following, we present you with several comparisons of elected politi These politicians have different personal characteristics and advocate for varying political goals. For each of these comparisons we would like to k what you think of them. Please take a close look at the descriptions of the two politicians before evaluating them in the next step.

Experimental treatments:

1) Name *[Christine Müller / Günther Schmidt / Yasmin Said / Omar Aslan]*
2) Political leaning *[leftist / centrist / rightist]*
3) Is currently in active politician at the [local / national / EU] level
4) Has recently been in the news *[to present a political proposal / for lying in a press conference / for abusing powers to favour family members / for taking br*
5) Seeks policy solution by *[bridging opposing ideological camps / standing up for the own ideological camp / responding to the majority of public opinion]*
6) Wants to *[maintain current measures to combat climate change / accelerate measures to combat climate change even if they are costly / slow down mea to combat climate change because they are too costly]*

1) **Which of these two politicians do you prefer?**
2) **Please rate how much you like each of the two politicians.**
3) **How well do you think each of these politicians represents you**

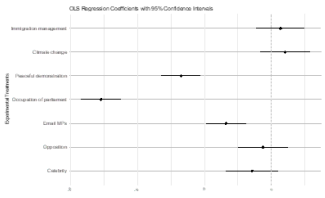
Key Lessons for Future Data Collection

1) **Careful with *Languages***
   a) Conduct multi-step language checks
   b) Be aware of the risk of measuring different concepts across languages
2) **Careful with *Fieldwork Agencies***
   a) Ensure they adhere strictly to all agreements
   b) Obtain written confirmation of all discussed issues
   c) Alternative: Consider programming the survey yourself
3) **Careful with Da*ta Protection***
   a) Ensure respondents cannot be identified
   b) If necessary, summarize survey categories after data collection

Key Lessons for Future Data Collection

1) **Develop fine-grained measurements to differentiate between related concepts of political disaffection** *(trust, confidence, legitimacy, representation)*
2) **Cover more types of affective issue polarization** *(international security, economic concerns)*
3) **Integrate other potential causes of political engagement in the vignette experiment** *(subjective motivations)*
4) **Over-time data collection**

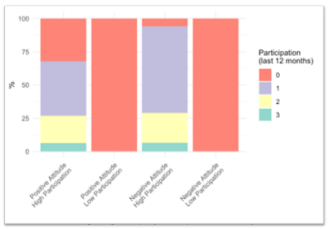
Research Application I



Note: base categories: "gender equality", "online petition" and "ordinary citizens".

Political participation
Data: ActEU survey
Survey section: Vignette experiment
Research question: What motivates people to politically engage?
Status: *Work in progress (Reinl et al. 2025)*

Research Applications II



Political disaffection

Data: ActEU focus groups + survey

Survey sections: political trust, democratic support & political participation

Research question: How can we capture political disaffection in Europe in a more holistic way?

Status: *Work in progress (Braun et al. 2025)*

Accessing ActEU data

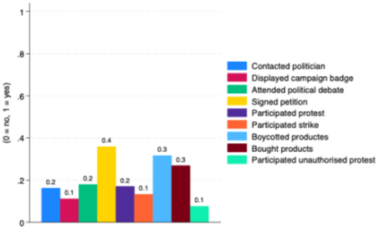
Scan QR code

- **Access via:** *GESIS Leibniz Institute for the Social Sciences*
- **Cost:** Download and use are FREE OF CHARGE
- **DOI:** https://doi.org/10.7802/2782

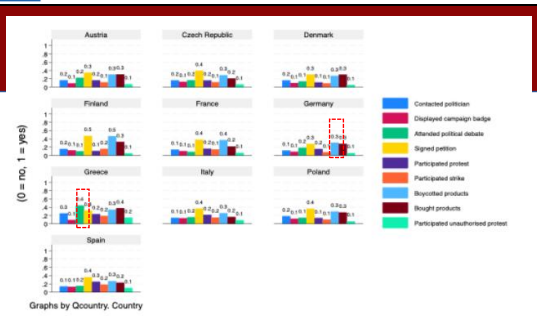
Hands-on Session

- Download the data set
- Browse through the data and make yourself familiar (15 minutes)
- Come together in small groups
- Figure out and discuss (15 minutes):
  1) Find the variables measuring non-electoral participation.
  2) Which type of non-electoral political action is most popular?
  3) Do we observe any interesting differences between countries?

Hands-on Session



Hands-on Session



Hands-on Session

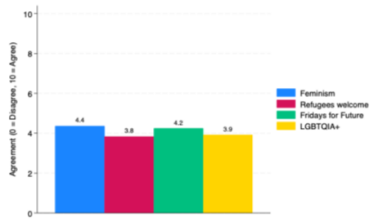
Figure out and discuss (15 minutes):

4) Find the variables measuring support for flags being displayed on the facades of local parliaments.

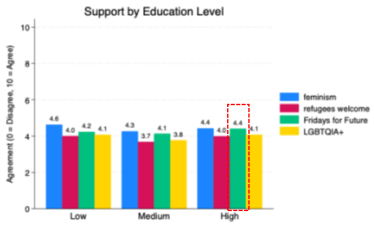
5) Which flags would most people support being displayed on the facades of their local and national parliaments?

6) Do we observe any interesting differences related to education?

Hands-on Session



Hands-on Session



Questions? Comments?

Thank you for listening

Appendix B: Presentation slides for the teaching script on web scraping

Summer Institute of Computational Social Science

Exploring Political Discourse through the Collection and Analysis of Digital Texts

What We'll Cover in this Workshop

- Digital Texts are Digital Traces
- The Good, the Bad and the Ugly of Working with Digital Texts (mostly from Social Media)
- Targeted Collection
- Collection Methods
- *Ab verbisper numeros*
- Hands-on Examples

Digital Texts are Digital Traces

- **Digital text is a form of digital trace** it wasn't created for research purposes.
- **Text as data methods** allow for **different research tasks** such as:
  - discovery (uncovering new conceptualizations),
  - measurement (quantifying concepts),
  - prediction (forecasting outcomes), and
  - causal inference (understanding relationships).
- **This workshop will focus on digital text** and its associated metadata, **collected** from social media, web sites and online news outlets.

The **Good** of Working with Digital Texts (mostly from Social Media)

A large part of society uses **search engines** and **social media** and by doing so they are leaving behind **(digital) traces** whose **exploitation** could open the door to:

- Better **understanding individual and social behavior.**
- Detecting and monitoring **epidemics.**
- Gaining deeper insight into **health and mental health** issues.
- Identifying the side effects of certain medications.
- Quantifying food consumption patterns.
- **Gauging public opinion.**
- **Predicting, for individual users,** sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, substance use, parental separation, age, and gender.
- Using social media as a **source for intelligence services** (aka SOCMINT).

The **Bad** of Working with Digital Texts (mostly from Social Media)

- **Social media is not representative of society** but instead displays **demographic biases** (even though sometimes the press interpret it as a barometer of public opinion).
- **Online behavior does not always reflect offline behavior** and is often blatantly **false**
- There is a **self-selection bias** whereby **the most extreme users** tend to be **the most active** making their views appear to be the majority.
- **Users** are well **aware** that **their collective actions** can **influence how an event is perceived** outside social media.
- **Bots and sockpuppets are widespread** in social media.
- There is a **content production bias** favoring what is recent, extreme, and unusual.

The **Bad** of Working with **Twitter/X**

- **Twitter has been the fruit fly of social media research** (Zeynep Tufekci).
- The **fruit fly** is a **model organism** used in all kinds of research on genetics and developmental biology. The **reasons** for this are **numerous from a biological standpoint, but many are a matter of convenience**
- Something similar happened with Twitter:
  - **it is not the most widely used social platform** and
  - it has **multiple issues** regarding how **representative** it is of the general population, but…
  - it **offered data access** that, compared to other platforms **made it the easiest medium** for researchers to study.

The **Ugly** of Working with Digital Texts (mostly from Social Media)

A large part of society uses search engines and social media, and by doing so they are leaving behind (digital) traces whose exploitation could open the door to:

- Better understanding individual and social behavior.
- Detecting and monitoring epidemics.
- Gaining deeper insight into health and mental health issues.
- Identifying the side effects of certain medications.
- **Quantifying food consumption patterns.**
- Gauging public opinion.
- **Predicting, for individual users, sexual orientation, ethnicity, religious and political views, personality traits intelligence, happiness, substance use, parental separation, age, and gender.**
- Using social media as a **source for intelligence services** (aka SOCMINT).

The **Ugly** of Working with Digital Texts (mostly from Social Media)

- **Doable and sensible are not synonyms.**
- The fact that **digital traces can be used to infer data or train models to make predictions does not mean it should be done.**
- Just because something has been **published** on social media does not mean it is **"public"** (in all senses of the term).
- **Saying the data was already public is not an excuse for causing harm.**
- It's also important to remember that **texts**—especially those from social media—are written by **people** and sometimes reveal sensitive, intimate, or vulnerable situations.
- In other words: **data is people!**

The **Ugly** of Working with Digital Texts (mostly from Social Media)

- For the most part, **digital traces are digital exhaust**—a byproduct never intended for direct consumption…
- In other words, they're not **"the new oil"**—bad enough; they're something worse: **"toxic waste".**
- Demographic, self-selection, and content production biases are minor issues in comparison.
- **Digital texts** also **carry** all the **prejudices and discriminatory, derogatory attitudes** that are constantly present in the online world.
- The **online world is an adversarial environment** where **malicious actors** and **false information distort the data and undermine the results and conclusions drawn from it.**
- Finally, we are in a **post-API era** marked by increasing **data inaccessibility.**

Targeted Collection

- Researchers usually employ **targeted strategies** to collect the data they deemed relevant for their research.
- This involves **focusing data collection** based on specific criteria:
  - Using **keywords, phrases or hashtags** to find relevant content within posts, articles, or blogs.
  - Identifying **data originating from or associated with specific actors** relevant to the research question.
  - **Selecting data sources** based on the specific online populations or communities being studied.

Targeted Collection: Keywords/Hashtags

- Collecting data using **keyword** or **hashtag** filtering is relatively **easy and** therefore **common** in the literature.
- However,
  - it may miss synonyms—i.e., there is a recall problem,
  - requires **continuous refinement** as terms evolve over time, and
  - often reflects the audience's point of view—the **"demand"**—rather than the "supply" side of the discourse.

Targeted Collection: Actor-Based Approaches

- Collecting data by focusing on **key actors**—such as politicians, news outlets, or NGOs—requires **substantial curation effort initially** but is easier to maintain over time.
- This approach mainly captures the **"supply"** side of political communication—unless audience reactions to these key actors are also included.
- Hence, it may be **biased toward elite perspectives** and risks overlooking discourse outside those selected actors.

Targeted Collection: Source Selection

- Targeting **particular platformsor communities** also requires a **curation** effort.
- The selection process can introduce **representation biases**—for example, in choosing which news outlets to include.
- **Manual identification can be challenging**—such as when selecting "fringe" political communities—and may require alternative approaches like **snowball sampling.**

Targeted Collection: Trade-offs and Biases

- Using keywords may lead to missing important discourse that uses different vocabulary or framing.
- Hashtag-based collection can overlook untagged content and may over-represent highly vocal subgroups.
- Any curation effort—whether involving keywords, hashtags, accounts, communities, or media sources—risks introducing representation biases.
- Depending on the approach, we may emphasize either the audience/demand side or the elite/supply side of political communication.
- **Therefore, wehave to be completely transparent about any decision informing data collection.**

Collection Methods

- We can **collect digital text**—and associated metadata—using **two main technical approaches:**
  - Programmatic access via **APIs**.
  - **Web crawling** and **scraping**—including both web and screen scraping.

Collection Methods: APIs

- Many digital platforms provide **Application Programming Interfaces (APIs)** that allow computational programs to access their data.
- These APIs act as web services designed for interacting with and **retrieving structured data.**
- APIs make it possible for researchers to collect data programmatically from external platforms like online social networks.

Collection Methods: APIs

- **Researchers** often need to **apply for developer access** and obtain specific credentials or API keys, agreeing to the platform's Terms of Service.
- They identify specific **"endpoints"** within the API documentation that correspond to the type of data needed (e.g., posts, users, comments).
- A **query is programmed**—typically using R or Python, specifying parameters like keywords, user IDs, time ranges, and the desired number of results.
- The **query is sent to the API** and the **platform sends back the requested data** typically in a structured format like **JSON**.

Collection Methods: APIs May Not Be Enough

- Some **platforms** simply **do not provide** a public API for data access.
- **Sometimes APIs return only a subset** of available data (e.g., recent content only, or sampled results).
- **Platforms may cap how much data you can retrieve** limiting large-scale research.
- Most APIs **don't provide full historical archives** making longitudinal analysis difficult.
- **APIs can change without notice** breaking your code or cutting off access entirely.

Collection Methods: Scraping as Last Resort

- Scraping lets researchers retrieve **visible content not included in APIs**
- It can **reach hard-to-access** communities or niche datasets.
- **Unavoidable when** platforms **do not offer APIs** or when **API access is limited or unsuitable** for the research goal.

Collection Methods: Scraping is Difficult

- Usually, **custom scrapers** need to be built for different websites due to variations in their structure.
- **Many sites detect and prevent automatic scrapers.**
- **Scraping effort is higher than using API access** and requires careful implementation to avoid overwhelming the target website's server—and to avoid detection.
- **Scraping raises legal and ethical questions** terms of service, copyright, and privacy must be considered, and balanced against the benefit of the research to be conducted.

A Little Bit about Text Preprocessing: Boilerplate Removal

- Natural language is extremely complex, and we have no intention of fully understanding it. The goal is to **apply a series of simplifications that transform something elaborate** like a press article or a social media post, **into data that can be handled by an algorithm.**
- To begin with, we will assume that **the starting material is digital text** proceeding from HTML pages, PDF documents, plain text stored in JSON, or similar formats.
- Depending on the type of source format, it may be necessary to **"clean" the text**, since not all content in the document will be of interest to us.
- For instance, if we are working with web content —either crawled or scraped —we will need to **remove "boilerplate"**.

A Little Bit about Text Preprocessing: Whitespace, punctuation, and capitalization

- Text is usually organized hierarchically and, depending on its length, it may or may not have chapters and sections, but it will always have paragraphs made up of sentences, which in turn consist of words.
- This structure is not especially relevant for our purposes, but it introduces a range of features in the original text that are best removed to simplify processing.
- For example:
  - Paragraphs are separated by a newline, which may be represented by one or two characters depending on the operating system (e.g., \n in Unix and \r\n in Windows).
  - Paragraphs end with a period.
  - The first word of a new paragraph always starts with a capital letter.

A Little Bit about Text Preprocessing: Tokenization

- On the other hand, in the languages we'll be working with most immediately (i.e., European languages), "whitespace" is used as a word separator, and you've likely already guessed that one foundational step for any automatic text analysis is to determine the frequency of each word.
- This might seem as simple as splitting the text using whitespace, but that would result in "words" like:
  - *hand,*—includes a comma at the end.
  - *(i.e.,*-includes a parenthesis at the beginning and two punctuation marks at the end.
  - *"whitespace"*—enclosed between quotations marks.
  - *word.*—includes a period at the end.
- You might think of replacing punctuation marks with spaces, but it's better to pause, recognize that this is a complex issue, and know that work has already been done in this area.
- We have access to tokenization algorithms that can perform this task–that is, splitting the text into its constituent tokens (words, digits, punctuation, spaces, etc.)

A Little Bit about Text Preprocessing: Terms

- Having introduced the concept of a **"token,"** it's important to present the concept of a **"term."**
- Terms are the first elements of interest that we want to access in our text "mining" process.
- Continuing with the mining analogy, the source text would be the ore, and the terms would be the metal to be obtained; terms (like metal) are not an end in themselves but are useful for a later purpose.
- Although it's true that in many cases a term will consist of a single token (e.g., the term "algorithm"), there will be times when we have **terms formed by multiple tokens** (e.g., the terms "climate change," "gender equality," or "far right"), the so-called multi-word expressions (MWEs).
- Attention!
  - For the purpose of this workshop, we will not worry about MWEs.
  - This is a "classical" approach to text processing; wordembeddings models may tokenize into full words or subwords.

A Little Bit about Text Preprocessing: Stopword Removal

- Stop words are those words that, despite frequent use, contribute little meaning to a text on their own.
- Removing stop words dates back to the 1950s and even though it's common practice it's not always a good idea.
- On another hand, what's a stop word? Let's take, for instance, *what*:
  - If it's a pronoun it can be treated as a stop word.
  - In "The Who" it would be a terrible idea to remove both *the* and *who*.
  - In "WHO declares COVID19 a pandemic" it is clear that its an acronym and not a stop word.
- That's why other method from natural language processing is also crucial during document preprocessing: in this case, part of speech tagging.

A Little Bit about Text Preprocessing: Stemming and Lemmatization

- **Stemming** consists of **trimming words** to their **lexical root or stem** not their lemma. For instance, universe, university, universitarian, universitarians.
- Stemming is an easy way to conflate semantically related terms. However, stemming algorithms are affected by overstemming, i.e., reducing unrelated terms to the same stem. For instance, universal, university and universe all conflate to the same stem: univers. Hence, it may be preferable—if possible—to apply lemmatization instead of stemming.
- **Lemmatization** is an NLP tool that performs a morphological analysis of the text to **identify the lemma** for each token—a lemma is a word that heads an article in a dictionary or encyclopedia. For instance, you won't find *is, are, am* or *was* in a dictionary, but *be*. A lemmatizer would conflate all those different terms into a single one: *be*.
- While stemming is a crude heuristic approach, lemmatization actually conflates all the inflected forms of a word into a single one. However, both methods are language dependent and, therefore, we need to know in which language the documents are written or automatically identify it —in Python we can use the package langdetect.

A Little Bit about Text Representation

- In **"classical" NLP and text mining texts** are often represented using **sparse lexical models** such as **bag-of-words** or **TF-IDF**.
- **Each unique word** in the document collection corresponds to a **different dimension** resulting in high-dimensional but sparse vectors —most words/dimensions in a document are zeroes.
- These models are called **lexical** because they rely on **exact word matches** —terms must appear verbatim to contribute to similarity.
- In contrast, **word embedding models** represent words (and sometimes entire documents) as **dense vectors in a lowerdimensional space.**
- By conflating dimensions (words) they **capture semantic relationships between terms**. This way, related words and semantically related documents have similar vector representations.

A Little Bit about Text Representation

- Let's look at the problem of using **sparse/lexical vectors** with an example:

Turkey is withdrawing forces from northern Iraq.

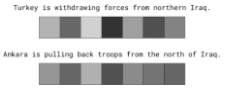
Ankara is pulling back troops from the north of Iraq.

- It is clear to anyone with a minimal knowledge of the events relating to Turkey and Iraqi Kurdistan that both texts are basically the same and yet their vectors in a sparse/lexical representation are very different...

ankara	forces	iraq	north	northern	pulling	troops	turkey	withdrawing
	x	x		x			x	x
x		x	x		x	x		

A Little Bit about Text Representation

- The image shows hypothetical **dense/semantic vectors** for each of the previous texts.
- Such representations, despite not being identical, reveal that **both texts are much more similar** than what a lexical similarity metric would indicate.
- Furthermore, users would interpret this "match" as evidence that **the model is capturing the underlying semantics** of the texts.



Ab verbis per numeros

- **Once we represent documents numerically** —either with sparse lexical models (e.g., bag-of-words, TF-IDF) or dense semantic vectors (e.g., embeddings)—**we can apply a wide range of techniques** such as:
  - Clustering (grouping similar documents),
  - topic modeling (discovering latent themes),
  - classification (assigning categories),
  - information retrieval,
  - automatic summarization,
  - and more...
- **Numerical representation is the gateway to scalable, automated text analysis.**

Hands-on Examples

- In this Google Colabnotebook we will:
  - Load and process annotated Twitter/X data on gender discourse.
  - Visualize issue salience and stance over time.
  - Apply Top2Vec for unsupervised topic modeling.
  - Use LLMs (Gemini) to generate humanreadable labels and explanations for the discovered topics.

Funded by the European Union

ActEU Activating European Citizens´ Trust in Times of Crises and Polarisation

So long, and thanks for all the fish..

Feel free to contact me at:

dani@uniovi.es

@PFCdgayo

@pfcdgayo.bsky.social

Funded by the European Union

ActEU Activating European Citizens´ Trust in Times of Crises and Polarisation

Appendix C: Presentation slides for the teaching script on focus group data

Methodological Mastery: Focus Groups– Design, Conduction, Analysis

Kristina Weissenbach
University of Duisburg-Essen
Kristina.Weissenbach@uni-due.de

From Puzzle to Focus Groups Results

1. Being puzzled about something: Unit of analysis & Research Question. The demand side of participation
2. Concepts & Conceptualization
3. Which kind of data for our research question? Why focus groups? One method of data conduction and analysis – or multiple?
4. Which design to triangulate our data?
5. *- Hands on: Reflect on your own research project & share it with the group -*
6. How to design, organize & conduct Focus Groups.
7. Ways of analyzing Focus Group data: What can you do with Focus Group data?
8. *Hands-on: Try out coding alongside the coding circle with ActEU focus group data. (30 minutes)*
9. Results: Efficacy as driver of participation?

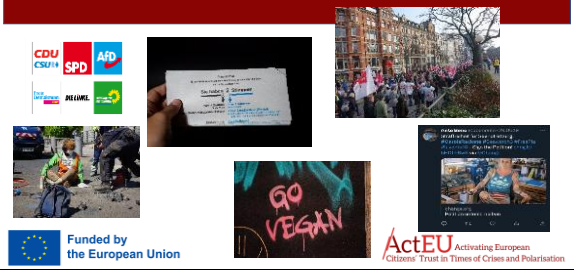
1. Being puzzled about something: Unit of analysis & Research Question. The demand side of participation

Political Participation

1. Which role does participation play for trust?

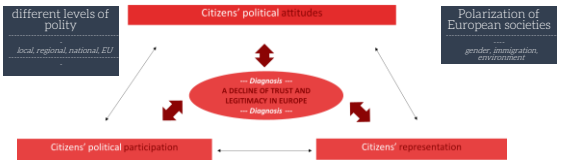
1.1 What kind of participation options do citizens demand for?
1.2 Are there differences between three main groups of citizens: disconnected – average – politically active citizens?

➢ *Please take a moment to take a first note: your main research question!*



2. Concepts & Conceptualization

- to explain or better understand your puzzle
- to generalize in a theoretical way
- for foresight research and scenario analysis and scenario building
- in our case:



Zooming into: Citizens political participation

Concepts:
Political Participation inside and outside the institutions of representative democracy
Conventional vs. unconventional forms of political participation
Collective vs. connective participation

➢ *Please take a moment to make your second note: which are your main theoretical concepts?*

3. Which kind of data for our research question? Why focus groups? One method of data conduction and analysis – or multiple?

- Focus Group Data
- Webscraping (Text as Data)
- Experimental Survey

➢ *Please take a moment to make your third note: New data or existing data? One method of data conduction or multiple? Which?*

Definition Focus Groups

Focus group research is "a way of collecting qualitative data, which—essentially—involves engaging a small number of people in an informal group discussion (or discussions), 'focused' around a particular topic or set of issues" (Wilkinson, 2004, p. 177)

- Focus groups environment is helpful for participants **to discuss perceptions, ideas, opinions, and thoughts** (Krueger & Casey, 2000).
- Focus group research is ideal to address process related "How"- questions rather than static "why" questions. "Understanding" rather than "explaining".
- Researchers have used focus groups for decades: In the 1920s, they were conducted to assist researchers **in identifying survey questions** (Morgan, 1998). In the early 1940s, Paul Lazarsfeld and Robert Merton, who are credited with formalizing the method of focus groups (Madriz, 2000), used focus group methods to conduct "a government-sponsored study to examine media effects on attitudes towards the involvement of the United States in World War II (Merton, 1987). These groundbreaking methodologists used focus group data to identify "salient dimensions of complex social stimuli as [a] precursor to further quantitative tests" (Lunt, 1996, p. 81).

- Lazarsfeld and Merton's research efforts constitute part of the legacy of using focus groups within qualitative research: (a) **capturing people's responses in real space and time in the context of face-to-face interactions** and (b) **strategically 'focusing' interview prompts based on themes** that are generated in these face-to-face interactions and that are considered particularly important to the researchers. (p. 899)
- Later, according to Greenbaum (1998), focus group data were collected and analyzed mainly for market researchers to assess consumers' attitudes and opinions. In the past 20 years, focus group research has been used to collect qualitative data by social science researchers (Madriz, 2000).

Benefits from using focus groups:

- economical, fast, and efficient method for obtaining data from multiple participants (Krueger & Casey, 2000), thereby potentially increasing the overall number of participants in a given qualitative study (Krueger, 2000).
- Another advantage to focus groups is the environment, which is socially oriented (Krueger, 2000).
- The sense of belonging to a group can increase the participants' sense of cohesiveness (Peters, 1993) and help them to feel safe to share information (Vaughn, Schumm, & Sinagub, 1996).
- Interactions that occur among the participants can yield important data (Morgan, 1988), can create the possibility for more spontaneous responses (Butler, 1996), and can provide a setting where the participants can discuss personal problems and provide possible solutions (Duggleby, 2005).

4. Which multi-method design to triangulate our data?

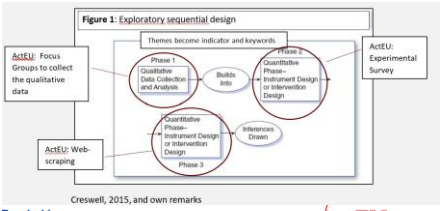
- Convergent parallel design
- Explanatory sequential design
- Exploratory and sequential design
- Embedded design
- Transformative design
- Mulitphase design

(Creswell-Plano 2011; Creswell 2013)

➢ *Please take a moment to a.) have a look at the Creswell/Plano-typology and b) make your fourth note* ***If*** *your research project is based on a multimethod design*

- which data to cover which aspect or dimension of your unit of analysis?

- Which multi-method design to triangulate your data?



- **Methodological value** of focus group discussions: instrument development function
- **Empirical value:** gain an in-depth understanding of the decline of citizens' approval and support of politicians and institutions in representative trust, demands and changes regarding participatory trust and the role of emotions in all trust dimensions.
- **Practical value:** analysis of focus group discussions enhance our understanding about *process-related, informal, emotional* aspects and the demand side of how different sub-groups of citizens of the European Union perceive their participation options and the way they feel represented.

5. Hands on!

Reflect on your own research project & share your sticky notes with the group on the flip chart!

6. How to design, organize and conduct focus groups

- between **1 and 2 hours** (Morgan, 1997; Vaughn et al., 1996)
- consist of **between 5 and 12 participants** (Baumgartner, Strong, & Hensley, 2002; Bernard, 1995; Johnson & Christensen, 2004; Krueger, 1988, 1994, 2000; Langford, Schoenfeld, & Izzo, 2002; Morgan, 1997; Onwuegbuzie, Jiao, & Bostick, 2004).
- The rationale for this range of focus group size stems from the goal that focus groups should include enough participants to yield diversity in information provided, yet they should not include too many participants because large groups can create an environment where participants do not feel **comfortable sharing their thoughts, opinions, beliefs, and experiences**
- Krueger (1994) has endorsed the use of very small focus groups, what he terms **"mini-focus groups"** (p. 17), which include 3 (Morgan, 1997) or 4 (Krueger, 1994) participants, when participants have **specialized knowledge and/or experiences** to discuss in the group.
- Keep **overrecruiting** by at least 20% of the total number of participants in mind!

Sampling & Recruiting

- Focus groups can be formed by **using preexisting groups** (e.g., colleagues at a place of work).
- Alternatively, these groups can represent **newly formed groups** that the researcher constructs by selecting members either randomly or, much more commonly, via one of the 19 or more purposive **sampling techniques** (e.g., homogeneous sampling, heterogenous sampling, maximum variation sampling, critical case sampling, or multistage purposeful sampling; Onwuegbuzie & Collins, 2007). Krueger (1994) and Morgan (1997) have suggested that **three to six different focus groups are adequate to reach data saturation and/ or theoretical saturation**.

ActEU Focus Group Sampling

Homogenous Target groups:

A group of **"disconnected"** citizens which is characterized by political disinterest, no participation, and social marginalization → **TG1**

The **"average"** group with mean levels of trust in the political system, without or with little political involvement. → **TG2**

The **"committed"** group which is actively involved in a.) traditional institutions of representative democracy (e.g. parties) and/or b.) in alternative modes of participation (e.g. demonstrations, protest, #activism). → **TG3**

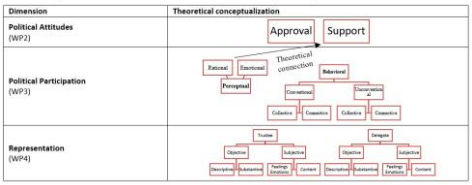
Heterogenous recruitment Criteria per target group and country:

All targets: Mix age, gender (except group 4, all female), residence, education, household income; all have the technical equipment to participate in an online focus group

Group of women: Mix of T1-3 (for target group allocation see country-specific criteria below)

- Company: Focus Groups run by IPSOS
- Field period: May – June 2023, 120 minutes
- Groups: 16 groups in four countries (Czechia, France, Germany, Greece)
- Screening & recruiting: 4 different groups (5-6 participants): disconnected citizens, average, engaged, women only

Focus Group Guide: From Concept to Conceptualization



Moderation Skills

- Ideal for the focus group to have a **moderator team**. This team typically comprises a moderator and an assistant moderator. The moderator is responsible for **facilitating the discussion**, prompting members to speak, requesting overly talkative members to let others talk, and encouraging all the members to participate.
- The moderator is responsible for taking notes that inform potential emergent questions to ask. In most cases, the moderator presents the focus group participants with a **series of questions**. However, instead, the moderator might present the members **with stimulus material** (e.g., newspaper article, video clip, audio clip) and ask them to respond to it.
- Alternatively, still, the moderator might ask the members to **engage in a specific activity** (e.g., team-building exercise, brainstorming exercise) and then provide reactions to it.
- In contrast, the **assistant moderator's responsibilities include recording the session (i.e., whether by audio- or videotape), taking notes, creating an environment that is conducive for group discussion** (e.g., dealing with latecomers, being sure everyone has a seat, arranging for refreshments), providing verification of data, and helping the researcher/moderator to analyze and/or interpret the focus group data (Krueger & Casey, 2000).

Moderation techniques in ActEU – How we captured trust Chinese Portraits

Chinese Portrait – Politicians: Now I would like to know more about your view of the German (French, Czech, Greek) politicians. In order for us to take a more playful approach here, I would not like to ask directly, but rather indirectly.

Think of German politicians at different levels. If they were a car, what car would they be?

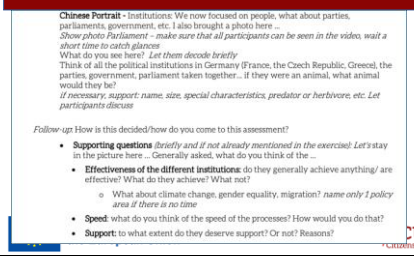
If necessary, support: brand, color, condition, age, special equipment, etc.
It is also allowed to name or invent different cars.

If necessary, ask the participants to first lump the politicians together; Let participants discuss

Follow-up- what are characteristics attached to/how do you arrive at this assessment?

Supporting questions *(briefly and if not already mentioned in the exercise): Let's* stay in the picture here ... Generally asked, what about the ...

- **Integrity**: honesty (don't pretend...)
- **Reliability** (do what they say...)
- What about the **sense of responsibility**? In whose interests are they acting? How do you come to this assessment?
- To what extent do they care about ordinary **people**?
- To what extent do they **deserve support**? Or not? Reasons?



Chinese Portrait - Institutions: We now focused on people, what about parties, parliaments, government, etc. I also brought a photo here ...
Show photo Parliament – make sure that all participants can be seen in the video, wait a short time to catch glances
What do you see here? *Let them decide briefly*
Think of all the political institutions in Germany (France, the Czech Republic, Greece), the parties, government, parliament taken together... if they were an animal, what animal would they be?
If necessary, support: name, size, special characteristics, predator or herbivore, etc. Let participants discuss

Follow-up How is this decided/how do you come to this assessment?

- **Supporting questions** *(briefly and if not already mentioned in the exercise): Let's* stay in the picture here ... Generally asked, what do you think of the ...
- **Effectiveness of the different institutions:** do they generally achieve anything/ are effective? What do they achieve? What not?
  - What about climate change, gender equality, migration? *name only 1 policy area if there is no time*
- **Speed**: what do you think of the speed of the processes? How would you do that?
- **Support:** to what extent do they deserve support? Or not? Reasons?

Picture shown in French Focus Groups

7. Ways of analyzing Focus Group data

- **There are many sources of focus group data, yet most researchers use only the actual text** (i.e., what each of the participants stated during the focus group) in their analyses. Multiple types of data can be collected during a focus group, including audiotapes of the participants from the focus groups, notes taken by the moderator and assistant moderator, and items recalled by the moderator and assistant moderator (Kruger, 1994). All of these data can be analyzed, yet they differ in the amount of time and rigor it will take to complete the analysis.
- **Transcript-based analysis represents the most rigorous and time-intensive mode of analyzing data**. This mode includes the transcription of videotapes and/or audiotapes, which commonly will result in 50 to 70 pages of text per focus group meeting.

What can you do with Focus Group data?

For example: Qualitative Content Analysis (QCA) and Thematic Analysis/Constant comparison analysis

....but much more: facial expression analysis, keywords-in-context analysis, narrative analysis, discourse analysis (for a review of analytical techniques see Leech&Onwuegbuzie 2007)

- There are different possibilities and levels of analyzing focus groups (Onwuegbuzie et al. 2009). In a first step, you can analyze the material via **qualitative content analysis (QCA)** following e.g. Kuckartz) (which we did in an abductive and iterative way to extract new ideas for the experimental survey and the web scraping)
- In short: "QCA is a method for systematically describing the meaning of qualitative material. It is done by classifying material as instances of the categories of a coding frame" (Schreier 2012, 1)

Qualitative Content Analysis

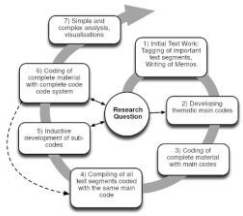
- Therefore, in **the center of the QCA are concepts, categories and the coding frame** "Categories are basic concepts of cognition; they are - generally speaking - a commonality between certain things: a term, a heading, a label that designates something similar under certain aspects" (Kuckartz 2019: 184). The codes can be developed in concept-driven (similar to deductive), data-driven (similar to inductive) or in a mixed manner (Kuckartz 2019:184f.).
- **After developing the first coding frame, coders assign text passages to the different categories The coding scheme will not be complete in the first round**. It is necessary to develop missing categories and be open to new emerging ones.
- However, instead of creating a theme from the codes (as with thematic/ constant comparison analysis), with classical content analysis, these codes then are placed into similar groupings and counted.
- **Intercoderreliability:** In order to ensure reliable results, it is important that several people code the material.

Thematic Analysis / Constant Comparison Analysis

- Thematic analysis is the process of **identifying patterns or themes** within qualitative data. Braun & Clarke (2006) suggest that it is the first qualitative method that should be learned as...it provides core skills that will be useful for conducting many other kinds of analysis' (p.78).
- It is a method rather than a methodology (Braun & Clarke 2006; Clarke & Braun, 2013). This means that, unlike many qualitative methodologies, it is not tied to a particular epistemological or theoretical perspective. This makes it a very flexible method, a considerable advantage given the diversity of work in learning and teaching.
- Goal of a thematic analysis is to identify themes (not numbers),i.e. patterns in the data that are important or interesting, and use these themes to address the research or say something about an issue. This is much more than simply summarizing the data;

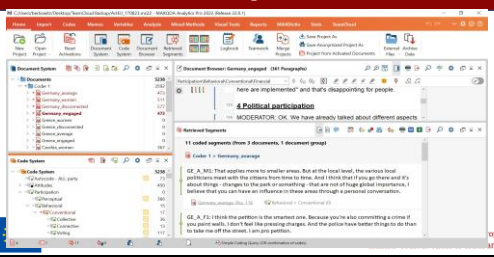
A good thematic analysis interprets and makes sense of it. A common pitfall is to use the main interview / focus groups questions as the themes (Clarke & Braun, 2013). Typically, this reflects the fact that the data have been summarized and organized, rather than analyzed.

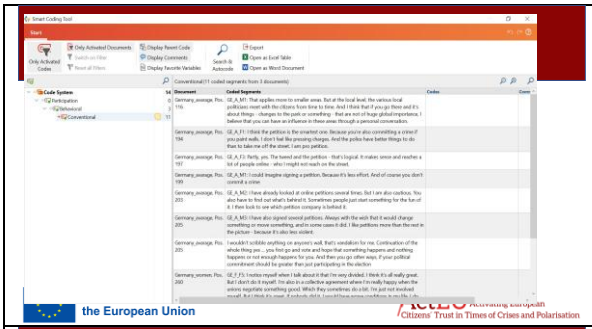
Coding-Circle in ActEu

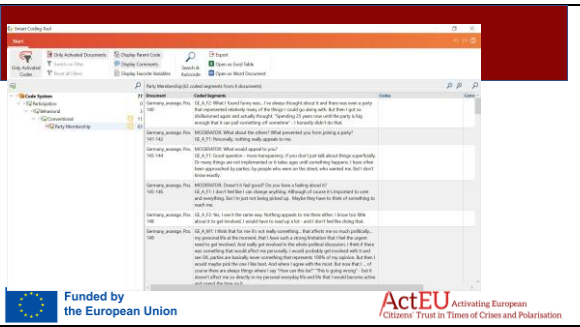


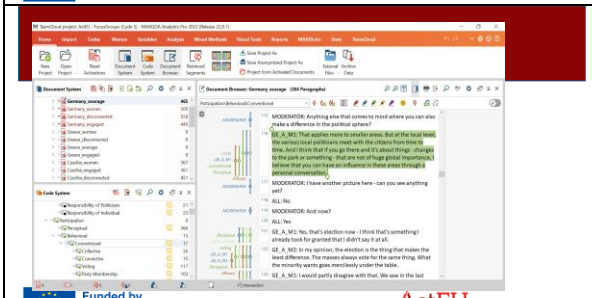
Kuckatz / McWhertor (2014)

Developing thematic main codes: Conventional Participation

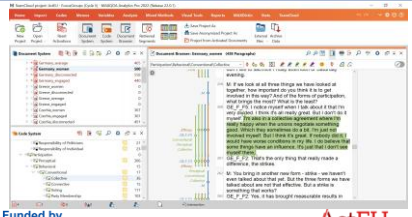




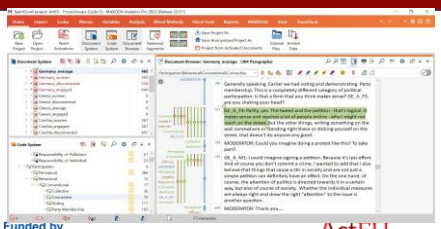




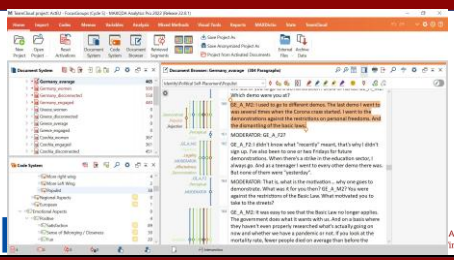
Collective participation



Conventional, Connective participation



Inductive development of sub-categories



Quick Q&A?

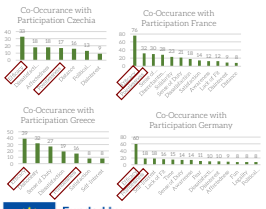
8. Hands-on!

Try out coding alongside the coding circle with ActEU focus group data!

Apply the coding-circle to one of the ActEU focus groups transcripts and to develop main and sub-codes in an abductive way:

- Tag important text segments according to the theoretical concept
- Develop thematic main codes according to the theoretical concept
- Code exemplary paragraphs of the ActEUfocus group transcript with main codes
- Develop thematic sub-codes inductively
- Complete the code system

9. Results: Efficacy as driver of participation? – Countries & Groups –



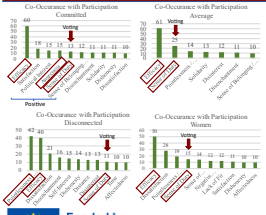
Belief in whether their actions will impact the desired outcome influences their decision to participate to a large extent

Do they receive a direct feedback for their involvement? Otherwise option is evaluated negatively

"F_F1: the vote, because if you have a politician who gets 52.3% of the vote, that's real representation, they'll be president. But if I go out to demonstrate, they get the 49.3 out, and that has no impact on everyday life" (France women).

Some positive notions in connection to local initiatives as well as voluntary work

"GE_A_M1: That applies more to smaller areas. But at the local level, the various local politicians meet with the citizens from time to time. And I think that if you go there and it's about things - changes to the park or something - that are not of huge global importance, I believe that you can have an influence in these areas through a personal conversation" (Germany average)



Pointlessness Disconnected

"There are countless ways! The best and easiest thing not to lose your sanity and your soul is, as GR_D_F3 said, non participation. Of course, there is a misconception that you compromise and accept what you are told, but there is nothing else you can do" (Greece disconnected)

Belief in whether their actions will impact the desired outcome influences their decision to participate to a large extent also in the groups.

Committed slightly more positive about the evaluation → e.g. awareness as influence more prominent

"GE_C_F2: To tie in with the topic of visibility I think they are all very important. Because visibility and attention are achieved through all of them. There is currently a lot of discussion about climate stickers. And the proportionality of the means of choice. Nonetheless, disobedience is mentioned again. But these are measures that create visibility and attention, which I think is important" (Germany committed)

Do they receive a direct feedback for their involvement?
→ True for different groups as well, less the committed

Going beyond counts! Going beyond verbal communication!

…..**counts should never be used to replace any qualitative data arising from focus** groups because by themselves they can present a misleading picture. In particular, the fact that the majority or even all of the focus group members express a particular viewpoint does not necessarily imply that this viewpoint is important or compelling. However, when contextualized, the use of counts can provide richer information than would be obtained by using qualitative data alone (cf. Sandelowski, 2001). Indeed, supplementing qualitative data with counts yields a form of mixed methods data analysis, or what is also known as mixed analysis (Onwuegbuzie & Teddlie, 2003; see also Morse, 2003). When used in this manner, enumerating the frequency of a particular viewpoint or experience actually expands the data set rather than reduces it (Onwuegbuzie et al. 2009)"

Going beyond counts => words in context

Going beyond verbal communication => nonverbal communication, e.g. facial expression analysis

Funded by the European Union

ActEU Activating European Citizens´ Trust in Times of Crises and Polarisation

Thank you for your attention!

Funded by the European Union

ActEU Activating European Citizens´ Trust in Times of Crises and Polarisation

About ActEU

How can we conceptualize and empirically measure political trust and legitimacy beyond the usual survey question "How much trust do you have in the parliament?"? Does the multi-level nature of European representative democracies require an identical level of citizen support at the regional, national and EU levels? How does social polarization on key policy issues of our times – immigration, climate change, and gender inequality– challenge the political trust in, and legitimacy of, democratic political systems? And what can policymakers and civil society do to master these challenges? ActEU aims at finding answers to these questions pursuing two overarching goals: In phase 1, we map and investigate persistent problems of declining trust, legitimacy and representation in Europe with a particular attention to the polarization of societies and the EU's multi-level structures. Providing an innovative conceptual framework on political attitudes, behavior and representation across Europe, we establish an original empirical infrastructure based on an innovative combination of methods and newly collected quantitative and qualitative empirical data (focus groups, experimental surveys, web scraping). In phase 2, these results will flow directly into the creation of a toolbox of remedial actions to enhance political trust in and legitimacy of European representative democracies. In cooperation with a newly created Civil Society Network, Youth Democracy Labs across 13 European cities and in exchange with political cartoonists "Cartooning for democracy", we will develop context-sensitive solutions for all polity levels and some of the most polarizing policy areas, and craft tailor-made toolkits for both policymakers and civil society and the educational sector. Finally, we deploy a differentiated dissemination strategy to maximize ActEU's scientific, policy and societal impact in activating European citizens' trust and working towards a new era of representative democracy.

www. acteu.org

ActEU Consortium

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.